The Journal *of* Parapsychology

Volume 2 SEPTEMBER, 1938 Number 3

EDITORIAL COMMENTS

THE American Psychological Association has planned a round-table discussion at the Columbus meetings in September, at which there will be an attempt, it is stated, to arrive at acceptable methods for the testing of the hypothesis of extra-sensory perception. To many persons this may appear to represent a certain amount of recognition by this psychological organization of the existence at least of the problem of ESP. But the limitation of the discussion to methods indicates a very tentative and guarded approach. Nevertheless, it must be recognized that the scheduling of this round table on ESP methods represents a step in advance, however slight it may be, and indicates a fair degree of tolerance for a radically informal problem on the part of the Association administration.

Readers of this JOURNAL may wonder, it is true, that four years after the publication of the Pearce-Pratt series there is still disagreement over methods for testing ESP. To our knowledge, no criticism has ever been raised against the methods of that piece of research. Conducted as it was with distance, independent records of cards and calls, even double witnessing of one series and rechecking of all records, there is probably no research more securely safeguarded in the entire field of experimental psychology. If we add to this the distance experiments of Riess, of Warner, and of MacFarland, with substantially similar results and conditions, there may be yet more wonder why we should still have to take up the threshold problems of methodology and conditions. On the other hand, the value of an opportunity to discuss methods and conditions with a new and a wider circle of scientific acquaintances must not be overlooked; and whether or not these conditions can be improved in respect of basic safeguards, there may well be variations of method suggested that will be of importance to ESP research.

A much more profound problem, however, than that of precautions is now confronting the investigators of extra-sensory perception. This too is a problem of methods and conditions, not, however, concerned with the safeguarding of tests against sensory perception. We refer to the conditions which will enable the elusive process of ESP to function measurably and more dependably; methods of treating the subjects and of providing mental situations which will evoke significant performance in the ESP tests. The desirable end of research into any new relation is the exercise of full control over it by manipulation of the conditions affecting it. We must presumably first discover these conditions by following up the successes and failures resulting from present explorations.

We have mentioned earlier that some investigators, indeed the majority of them, have apparently been successful in repeating the tests for ESP with significant deviations. But a considerable number have not been successful. Even in some instances in which considerable range of subjects was included in the test, failure to obtain significant deviations still resulted. Since other work indicates a fairly wide distribution of ESP ability, it seems safe to conclude, then, that there are experimenters who can obtain extra-chance results in conducting ESP tests, and those who simply cannot; at least, with our present knowledge of how to go about them.

The general problem of what is favorable and what is unfavorable to ESP performance has of course been a dominant one from the beginning of the research, and much has been learned about it. In the earlier Duke experiments there was, for example, unmistakable evidence that sodium amytal in a certain dosage was unfavorable. Again, various experiments have indicated since that certain test procedures are for certain subjects much more favorable than others. This was shown in Woodruff and George's comparison of certain matching and calling procedures. In this comparison subjects differed likewise in response to the interposition of screens, as they did also for Carpenter and Phalen.

Marked differences in success for a number of subjects were found by Sharp and Clark when the same subjects were tested both in groups and individually. Likewise, the position of the experimenter in the room (whether facing or behind the pupils) appeared to affect the subjects in Bond's tests of retarded school children. Bender found important differences in the position of the stimulus card. Tyrrell obtained temporary but marked differences in scoring with the introduction of new modificitions in his apparatus, as had Rhine earlier with new

changes in test procedure. These instances by no means exhaust the range of variations in results corresponding to variation in conditions of the test; and much more work of this character remains to be published.

The present number of the JOURNAL adds further to our information regarding the effect upon scoring of varying conditions in the tests. Perhaps the nicest experimental control of this variation occurred in the experiments of Stuart, studying the effect on scoring of the rate of responses as governed by a metronome set at different rates. He found that extra-chance scoring was definitely restricted to a certain range of rate of response corresponding to the favorite tempo of the subject. Hardly less definite and clear-cut, however, is the distinction between score levels in the two techniques compared in the experimental work reported by MacFarland, the comparison of the clairvoyant DT technique with that of GESP. While both procedures gave statistically significant results, the difference is very marked and is quite significant in itself.

That good ESP subjects can apparently be driven even into significantly negative deviations under certain test conditions appears to be demonstrated in the work reported by Rhine with cards enclosed in sealed opaque envelopes. With sufficient variation in delay in checking, the subjects apparently can be made to swing all the way from significantly positive deviations to negative ones.

But what stands out most among the conditions favoring and disfavoring ESP performance is the influence of the experimenter (or other persons) present during the test; and this is most definitely indicated in the work reported in the present number.

It may be recalled that in the earlier Duke work the presence of visitors was reported by Rhine[1] to have an inhibitory effect upon scoring, bringing it down to the level of chance until a certain period of adjustment had passed, after which the scoring would rise to its usual level. This and other indications served to call attention to the influence of those present during the test.

As mentioned above, and as is well known, there have been experimenters who have not been able to obtain extra-chance scores from their subjects, although the number of subjects tested would apparently warrant the supposition that some such ability (if it be distributed through our population in the manner indicated by observation) must have been possessed as a potentiality by several members of the group. It would appear, then, an entirely reasonable hypothesis to hold that

[1] Extra-Sensory Perception. Boston: Bruce Humphries, 1934, pp. 75 ff.

the presence or the activity of these investigators inhibited the performance of the subject.

The hypothesis that some experimenters might have an inhibitory effect upon subjects taking the ESP tests under them (or it may be preferred to say that certain *other* experimenters possess the capacity to evoke successful ESP performance) was given its first clear-cut objective demonstration by Price and Pratt, although earlier observations (e.g., Clark and Sharp) had given indications of this effect. Either that Pratt's influence was inhibitory or that Price's was productive of ESP scoring seems to be a necessary conclusion from their report.

Continuing the consideration of this hypothesis in the work of the present number of the JOURNAL, we might point in passing to the preliminary study by Drake of the feeble-minded boy whose possible ESP responses seemed to be able to function only with his mother acting as agent or sender.

More specifically following up the Price-Pratt treatment, however, is the work of MacFarland. In this instance, too, there was one experimenter who had not in earlier tests been successful in obtaining extra-chance averages from his subjects, and another who had. Here again the two experimenters conducted a joint experiment; but in this instance they were not actually in the presence of the subject, as were Price and Pratt, but co-operated at the other end of a signal arrangement, with the subject knowing that there were two experimenters sending independently. Even so, the previously inhibitory experimenter still was unable to get appreciably above chance, while the same subjects scored significantly extra-chance for the experimenter who had usually been successful in previous work. In addition, then, to confirming the Price-Pratt observation regarding the subject-experimenter relationship, this work suggests that (at least where subjects and experimenters are already well acquainted) it is not even necessary for the experimenter to be immediately present during the test in order to exert the inhibitory or the helpful effect which is demonstrated in the present results.

It seems clear that there is opened up here a most important group of problems, and above all it will be necessary to avoid hasty conclusions in this direction. In so far as the effect on scoring is concerned, favorable or unfavorable, it is clear, we think, that the experimenter himself has some influence. It seems desirable, however, to reserve judgment as to just how far this influence may be exerted indirectly and *in absentia* (i.e., it would likely be different using new subjects as Pratt and Price did). At the same time, we must be open to certain definite possibilities that may be far-reaching.

Above all it is in order to suspend judgment in weighing the consequences and bearing of the results obtained by those repeating the ESP experiments and obtaining only chance results. As was suggested almost a year ago in the Pratt and Stuart *Handbook for Testing Extra-Sensory Perception,* the profitable thing to do for an experimenter who has not had success in obtaining high scores is to secure the help of someone who can succeed. This is common practice in other professional relations in which the personality of the practitioner concerned is inhibitory.

On another point, too, there should be easy general agreement: Regardless of the above actual experimental results, it should be recognized that the mere possibility that the experimenter himself may be a factor in the determination of the test results is enough to defeat the (already dubious) argument that negative results prove anything about ESP. That is, it should be clear to all that failure on the part of certain experimenters to confirm the ESP hypothesis carries no logical weight against the many confirmations obtained by other experimenters.

The fallacy of regarding a *contrary* finding as *contradictory* is exemplified in a recent discussion by C. E. M. Joal, an English philosopher. In an article on psychical research in the July *Harpers* he dismissed the ESP research with the amazing reasoning that two fellow Englishmen had repeated the tests and had gotten only chance results! It can scarcely be believed that the confirmation of a dozen American college laboratories, involving a good many more than a dozen experimenters, was entirely unknown to the writer.

* * *

The JOURNAL will publish brief reports under "Research Notes" of ESP researches which have yielded only chance results, as these become available. It is desirable, at least for a time, to keep these in the total picture for what they may contribute to the problem of discovering conditions determining success and failure in ESP tests.

* * *

The explorations in ESP keep raising concurrently new problems and new applications of statistics. Fortunately, mathematicians are cooperative in meeting these problems. The problems raised by Leuba in this issue are part of a general series of sampling problems needing consideration. A substantial empirical attack by Greenwood on some of those more immediately concerning ESP appears also in this number. These problems are not confined to ESP research but are general ones. Though their solution may be crucial only to results of marginal significance, there is a recognized need to evolve a more specific and elaborate code of procedure for the use of statistical methods.

DISCRIMINATION SHOWN BETWEEN EXPERIMENTERS BY SUBJECTS

JAMES D. MacFARLAND*

ABSTRACT: Five selected subjects made 15,300 calls equally divided between the GESP and DT test procedures. Each call was aimed at two target decks and counted twice, making 30,600 trials. An average of 5.48 was obtained, and a critical ratio of 8.37. Two experimenters, one with each target deck, were present throughout; all recording was doubly witnessed, and the results were independently rechecked. Sensory cues were eliminated by having the subjects located several rooms away from the cards and experimenters, with a signalling device used by the subject to communicate his calls to the experimenter.

In the research here reported, two experimenters were compared, one of whom had been consistently "successful" in previous work and the other "unsuccessful." The subjects and the external conditions were the same for both. But the two experimenters still continued at approximately their previous levels of success, one at chance and the other consistently above.

The GESP procedure gave higher score averages than DT for all five subjects. Notably higher scoring was achieved in GESP when it happened that both experimenters were looking at the same card. No U-curves in this adaptation of the DT procedure were discovered, contrary to earlier experience with DT.

Mr. MacFarland is Instructor in Psychology at Tarkio College and has contributed considerably to earlier publications in this Journal.

IN RECENT months, in an attempt to explain the failure of some investigators to obtain extra-chance scores in ESP tests, some attention has been given to the subject-experimenter relationship.

Sharp and Clark,[1] classifying four experimenters in their attitudes toward ESP as "definitely skeptical," "not committed," "favorable" and "very favorable," found that averages in the same respective order for a varying number of trials were 4.30, 4.86, 5.88 (one subject only) and 5.36. They tentatively concluded that "on the basis of [the data] reported, there seems to be some evidence that the observer in experiments is more than a mere passive recorder. His attitude and actions would seem to be to some extent mirrored in the scores of his subjects."

* The co-operation of my colleague, Mr. Dale Hutchinson, has made this study possible. I am also greatly indebted to Dr. R. W. George for his generous help.

[1] "Group Tests for Extra-Sensory Perception," J. PARAPSYCHOL., I, 136.

But the question may well arise as to whether the mere possession of a favorable attitude is sufficient to insure success.

In the preceding issue of the *Journal of Parapsychology* Pratt and Price published a report of research dealing specifically with the problem of the "successful" and the "unsuccessful" experimenter. Working with similar groups of subjects and with similar techniques, one of two experimenters obtained scores of considerable extra-chance value, while the other obtained only results which were easily within the range of chance. Both experimenters concerned were entirely open-minded toward the research. The crucial point in the experiment came with the next step. The two experimenters began operating as a team, both handling the cards, checking scores, observing, and so forth. After a half-way stage which was unsuccessful, an attempt was made to approximate as closely as possible the experimental atmosphere maintained previously by the "successful" experimenter. The scoring rates of subjects then rose to the average maintained previously by the "successful" experimenter alone.

The present research attacks another phase of the same general problem of the personal influence of the experimenter. Two experimenters are again employed for the purpose of comparison. Experimenter J. M. has been consistently successful in obtaining extra-chance score averages from subjects.[2] Work done by experimenter D. H. has been approximately chance. Both experimenters are favorable to the ESP hypothesis. In the present research a set-up was employed whereby the two experimenters worked simultaneously with the same subjects, each experimenter with his own deck of ESP cards. Subjects, however, worked at a distance from the experimenters so that there was no contact throughout the experimental period except that possible by means of the signalling device used at the beginning and end of the run. The problem was to determine whether under such conditions the subjects would consistently hit more cards in the deck handled by one experimenter than in that of the other; that is, whether the relative success and failure of the two experimenters would result in the joint experiment as it had in their earlier individual research.

It was also regarded as important to discover further whether any differences found would persist through test procedures permitting both telepathy and clairvoyance and those permitting only clairvoyance. Presumably a more personal role on the part of the experimenter might be

[2] MacFarland and George, "Extra-Sensory Perception of Normal and Distorted Symbols," J. PARAPSYCHOL., I, 93; and other work included in reports by others.

expected with telepathy than without it. To make this further exploration the GESP and DT procedures were used in equal proportion.

PROCEDURE AND CONDITIONS

Apparatus. The experimental set-up was designed to eliminate the possibility of sensory cues and recording errors. To this end an electrical signalling device was constructed whereby subjects made calls two floors distant from the ESP cards and the experimenters. The device itself was simple. The receiving part of the apparatus, located in the experimenters' office, consisted of a small platform bearing five small compartments each of which contained a light and each of which was marked by one of the five ESP symbols. Also located on the platform were a push button to be used in lighting a small light on the subject's part of the apparatus and a small buzzer by means of which the subject could indicate his readiness to begin the experiment. The experimenters' device was located in the third floor office of the administration building at Tarkio College.

In the subject's office located on the first floor of the same building was a small panel bearing five telegraph keys, each identified with one of the ESP symbols. As the subject pressed a key a light flashed in the appropriate compartment in the experimenters' set. Also located on the subject's panel was the small light by which the experimenters indicated their readiness to begin, and a push button to be used in sounding the buzzer mentioned above.

Procedure. Two standard ESP techniques were used. In the GESP technique, experimenters D. H. and J. M. sat side by side in front of the receiving set. Each experimenter held a pack of cards which he shuffled thoroughly. The button on the platform was pushed three times to give the subject his cue to begin calling. Each experimenter looked at the top card of his deck momentarily, placed the card face down on the desk on which the set was located, and each recorded independently the call made by the subject as it was indicated by the lighting of one of the compartments. This continued through a standard run of twenty-five cards, at which time the experimenters signalled with their push button that the run was completed. No other signalling occurred; the pace was set by the subject. The double record of the subject's calls secured by the independent recording of the two experimenters was checked by placing the columns parallel to each other and both experimenters inspecting each record. Both experimenters then checked the two decks of cards against the doubly-recorded order of calls. In this way there was secured double witnessing of all checking.

And in view of the continual comparison of the results of the two experimenters, more than ordinary vigilance may be safely assumed.

The DT technique was used with the same precautions. In this technique, however, the experimenters did not disturb their decks of ESP cards till after the completion of the run of twenty-five cards. The difference in the two methods, then, lies in the fact that in the GESP test both clairvoyance and telepathy might have been responsible for the results while the DT technique was designed as a measure of clairvoyance only.

The GESP procedure was followed exclusively during the first half of the experiment, and then DT was used in unbroken sequence. This fact must be taken into account in the comparison of techniques. As a further precaution, in order to eliminate from the results all possible errors in recording, an independent check of the records for the entire experiment was made by another investigator.[3] This included also a sample of 100 runs checked to ascertain if any symbol did not occur five times in each run. This is a guarantee of accurate recording where the cards are recorded from an inverted pack and only one card exposed at a time (thus avoiding danger of inversion of two cards exposed). There were no errors discovered.

Instructions. The subjects knew from the beginning that there were two experimenters, J. M. and D. H. Likewise, they understood that two decks of cards were being used simultaneously and that a hit would be credited for either one or both. They were told to call until the stop signal was given.

On the other hand, the subjects did not know that there were two techniques being used, nor did they know the exact working of either. From post-experimental questioning[4] it seems likely that the subjects had the impression that something like the GESP technique was used. This is understandable, inasmuch as they had previously run many trials of GESP and few of DT prior to the time of the present experiment.

Subjects. The five subjects were selected on the basis of previous performance in other experiments. They were all men students at Tarkio College. They received no pay for their services, and their generous help is gratefully acknowledged.

RESULTS

Grand totals. The results obtained from the entire experiment, irrespective of experimenter and procedure, give a total of 1,224 runs

[3] Mr. J. L. Woodruff, of Duke University, whose assistance in this as in other tedious work on these data is gratefully acknowledged.

[4] By Mr. Woodruff.

(i.e., 30,600 single trials) with a total positive deviation of 596, an average number of hits per run of 5.48. This yields a critical ratio of 8.37. The individual performance of the five subjects for both conditions and both experimenters may be compared in Table I.

TABLE I

TOTAL RESULTS OF INDIVIDUAL SUBJECTS

Subject	Runs	Deviation	Average	C. R.
V. H.	444	+251	5.57	5.83
W. S.	360	+169	5.47	4.36
W. B.	80	+ 31	5.39	1.69
H. H.	152	+100	5.66	3.97
D. E.	188	+ 45	5.24	1.60
Total	1,224	+596	5.48	8.35

Comparison of experimenters. Prior to the present experiment D. H. and J. M. had acted as experimenters for a great number of trials on a project not yet reported. Almost without exception the GESP technique was employed. As a basis for the terms "unsuccessful" and "successful" as applied to the experimenters, all of D. H.'s previous GESP work, 397 runs, and a like amount of J. M.'s GESP work, have been compared. The work by D. H. falls roughly into periods, and that of J. M. has been selected on a basis of these periods. The scores themselves were in no way taken into consideration in this selection.

Table II gives the results of the comparison of prior and present GESP work by the two experimenters.

TABLE II

COMPARISON OF EXPERIMENTERS D. H. AND J. M., INDIVIDUAL AND JOINT EXPERIMENTS

	Individual				Joint			
Experimenters	Runs	Dev.	Av.	C. R.	Runs	Dev.	Av.	C. R.
D. H.	397	+ 3	5.00	0.07	612	+ 29	5.09	0.57
J. M.	397	+460	6.15	11.31	612	+567	5.92	11.00

In the present research, experimenter J. M. conducted 612 runs and obtained a positive deviation of 567, which gives an average of 5.92 and a critical ratio of 11.0. On the other hand, experimenter D. H. in the same number of runs had a positive deviation of 29, which yields

an average of 5.09 and a critical ratio of 0.57 for identical conditions and the same subjects.

Comparison of procedures. The 612 runs of GESP give a total positive deviation of 438 for an average of 5.71 and a critical ratio of 8.6. Compared with this 612 runs of DT give a positive deviation of 158 for an average of 5.25 and a critical ratio of 3.13. The critical ratio of the difference is 3.9.

The drop in scoring in the change from GESP to DT is more closely shown when we examine the results in detail, and for this purpose Table III is presented with individual averages for both procedures and

TABLE III

COMPARISON OF PROCEDURES GESP AND DT, SHOWING INDIVIDUAL SUBJECTS' RESULTS

GESP TECHNIQUE

Subject	D. H. as Experimenter				J. M. as Experimenter			
	Runs	Dev.	Av.	C. R.	Runs	Dev.	Av.	C. R.
V. H.	111	3	5.02	0.14	111	172	6.54	8.00
W. S.	90	5	5.05	0.26	90	114	6.26	5.89
W. B.	20	— 2	4.90	0.22	20	19	5.95	2.07
H. H.	38	17	5.45	1.35	38	42	6.10	3.34
D. E.	47	21	5.44	1.50	47	47	5.76	2.57
Total	306	44	5.14	1.21	306	394	6.28	11.10

DT TECHNIQUE

Subject	D. H. as Experimenter				J. M. as Experimenter			
	Runs	Dev.	Av.	C. R.	Runs	Dev.	Av.	C. R.
V. H.	111	— 18	4.83	0.83	111	94	5.84	4.32
W. S.	90	13	5.14	0.66	90	37	5.41	1.90
W. B.	20	3	5.15	0.32	20	11	5.54	1.20
H. H.	38	19	5.50	1.51	38	22	5.57	1.75
D. E.	47	— 32	4.31	2.29	47	9	5.18	0.60
Total	306	— 15	4.94	0.42	306	173	5.56	4.84
Grand Total	612	29	5.09	0.57	612	567	5.92	11.00

for both experimenters. In this table it may be seen that with experimenter J. M., for whom all of the significant scoring was done, there was not only a large drop in average from GESP to DT (6.28 to 5.56), but the averages of every one of the five subjects definitely dropped with the change.

Correct coincidence in two packs. The subjects' calls were matched against two independent card orders. The question arises as to the relative frequency of hits when cards in the independent decks corresponded (which should occur on an average of five times per run) as compared with the frequency of hits when the cards of the two decks did not correspond. Or, stated in other terms, in the GESP techniques when J. M. and D. H. were visualizing the same symbol was that symbol hit more frequently than in the cases when they were visualizing different symbols?

Chance should give 1,530 card coincidences for the 306 runs of GESP for each of the two experimenters, and of these a total of 306 hits would be the number most expected by chance; the same number for the DT technique when coincidences occurred.

It is obvious, however, that should the cards of the two independent decks coincide more frequently than expected by chance, there would be greater opportunity for hits by the subjects' calls, and high scores would be merely an artifact of a fault elsewhere in the experiment, probably in the shuffling of the two decks. A check was therefore made in which the independent card orders in the GESP work were matched together; a negative deviation of seven (i.e., a total of 1,523 as against 1,530 expected) was obtained. In other words, there was obtained the expected average of five coincidences per 25 between the two decks of cards. We can then safely proceed to evaluate the deviation of hits when the cards coincided and compare these with the general run of hits.

There were 1,523 card-coincidences in GESP between the decks of D. H. and J. M., and of these 1/5 or 305 would be expected (most probably) to be hit by chance alone. The actual score total is 428. This is 123 above expectation and gives an average per 25 of 7.02, to be compared to that of 6.28 with J. M. and 5.14 with D. H. for the total calls made.

The corresponding figures for the DT work are: Hits of corresponding cards (of the two DT decks), 332; the average per 25 is 5.43, compared to 5.56 with J. M., and 4.94 with D. H.

It is interesting to note also that the average for the total hits made on card-coincidences for both subdivisions (GESP and DT) is 6.21, whereas that of the total trials on all cards is 5.48. That for J. M. alone is 5.92.

The DT curve. Rhine, Pegram, and Gibson have noted in published reports the occurrence of a phenomenon now known as the DT curve. When the DT technique was used, their subjects hit more frequently the first and last five cards than the three middle groups of five cards

down through a pack. When plotted, the number of hits yielded a U-shaped curve.

In the present case, curves have been plotted for both experimenters for both the GESP and DT techniques. However, the phenonmenon of the U-curve for DT is missing, as the results below make obvious. There is a notable advance in score level toward the end of the run in J. M.'s work, both DT and GESP. It should be kept in mind that the conditions in these experiments differ from those referred to above.

TABLE IV

AVERAGE NUMBER OF HITS PER 25 TRIALS FOR GROUPS OF FIVE DOWN THROUGH THE RUN

	GESP					DT				
	1st Five	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
J. M.....	5.65	6.37	6.38	6.07	6.96	4.99	5.57	5.42	5.78	6.09
D. H.....	5.21	4.99	5.11	5.22	5.23	5.04	5.16	4.93	4.99	4.62

DISCUSSION

Grand totals. An average of 5.48 and a critical ratio of 8.37 for 1,224 runs place the total results definitely in an extra-chance category. What are the alternative explanations of this extra-chance scoring?

1. The extra-chance deviation may be the result of sensory cues communicated in some way to the subjects.

2. The extra-chance deviation may be a result of intentionally or unintentionally inaccurate checking of results.

3. The extra-chance deviation may be a result of ESP.

The first alternative seems obviously untenable. Two floors' distance and a number of intervening walls definitely rule out a suggestion that subjects obtained sensory cues from the cards. The same objections hold for a recently popularized suggestion that the experimenters may have communicated by unconscious whispering. A deliberate shout could not have been heard under the conditions of the experimental set-up.

The only communication between subject and experimenters during the course of a run was *from subject to experimenter*. Thus there is no likelihood that the experimenters influenced a subject's calls by a varying method of signalling.

In considering the second alternative, it is recognized that if the strictest precautions had not been taken, such an explanation would be

a dangerous possibility. The most obvious danger lies in the fact that the experimenters in looking first at the card and then at the signalling apparatus might write down the card looked at rather than the signal given. Is it possible that this was done? The checking together of the two sets of records made by each experimenter independently overcomes this objection. In the few cases where a difference was noted in the records no hit was scored as correct. Likewise, J. M. obtained very definite extra-chance scores by the use of the DT technique where cards were not looked at until the end of the run. In all cases when checking was necessary it was by the very nature of the experimental set-up doubly checked.

Added to the above precautions, in order to insure absolute accuracy, there was as stated above a recheck made of all data by an independent investigator. The precautions noted above seem ample safeguards even against intentional errors.

There appears to be left only the alternative of ESP as an explanation of the extra-chance scores. This hypothesis is strengthened further by the fact that the results gained in this experiment are entirely consistent with results gained by the same experimenters in other projects.

The two experimenters played identical roles in these tests, in so far as overt action goes. Yet very different results are found for the two. If the GESP work alone is considered, it is possible to account for the difference by supposing J. M. merely a better sender. But this will not account for the DT results. It is left to suppose only that some personal relation between the subjects and J. M. as experimenter is more favorable to the success in the tests.

Whether or not the advantage of J. M. was due to his staff position and his greater responsibility in the research is not possible to ascertain. The fact that in earlier work[5] he had been equally successful when his position was correspondingly the same as D. H.'s in this report, does not favor this view.

On the other hand, it is difficult to point at present to personal differences adequate to account for the results. Some subjects expressed a preference for J. M., but this may have been on account of earlier success with him. Also, they stated that they received clearer impressions (images) of the cards when J. M. was sending, but it cannot be concluded that this was due to superior ability as a sender on J. M.'s part. It is possibly due to the greater confidence the subjects had in him. But the simple fact stands out, unexplained as yet though it be,

[5] MacFarland and George, "Extra-Sensory Perception of Distorted Symbols," J. PARAPSYCHOL., I, 93.

that as a result of earlier success the different personal relations of subject and experimenter made a difference. It is at least demonstrated, then, that a mere desire to succeed is not enough to make a successful experimenter in ESP work.

Comparison of techniques. A glance at the comparative averages shows that the GESP technique appeared to be superior to the DT method for the obtaining of high scores. The question arises whether this is due (1) to the telepathic feature, (2) to the fact that the GESP tests use a "card by card" technique whereas the DT depends on the deck being unbroken till the end of the run, or (3) to a decline with mere length of series, the DT coming after the GESP. The present research does not make this comparison in its most ideal fashion, but is submitted for the indicative value it has on the still unsolved problems of method comparisons.

It is submitted that the lower DT scores are not due to a general decline since there is evidence of a decline neither during the GESP half nor during the DT half of the research.

As to the other two alternatives, there is no certain answer at present, though it seems likely that it is the absence of the "card by card" procedure rather than the exclusion of telepathy which is responsible for the drop with DT. In general, earlier reports have shown as great differences between BT (card by card) and DT as those given here between GESP and DT. Also, no general superiority of GESP over BT has yet been established. But this problem awaits further work.

The DT curve. It seems possible that the fact of the isolation of the subject from record sheets and all criteria which would help him keep track of the place of the trial in the run may be responsible for the absence of the usual U-curve reported by others. It is found in other work conducted by J. M. to be reported later that in tests involving the subjects' keeping their own call records on regular record sheets, the U-curve appears in the DT data. This strongly suggests that the method of response in the DT reported above may be responsible for the absence of the U-curve.

Higher average scores on card-coincidences. It appears from the results that the subject received an advantage from having both the experimenters look at the same symbol. Success was at least greater than for J. M.'s deck alone. This suggests enhanced scoring with two senders, even if one is not ordinarily successful. This problem too is under further investigation.[6]

[6] If it be inquired whether the greater success with card-coincidences than without might not result from error in recording the cards, making more than the normal number of coincidences, this is checked by the Woodruff recheck showing

Another interesting problem arises in the fact that experimenter D. H.'s scores for GESP are above chance to exactly the extent of the number of hits made on the card-coincidences over and above the average of J. M. alone. This suggests that D. H.'s deviation was given entirely by the hits made when his card coincided with J. M.'s. Yet it must be recalled that the hits for these coincidences were the highest made, higher than for J. M.'s deck alone; hence D. H. must be credited with a role in these hits of coinciding cards. The question arises as to what enables D. H. to function as a sender at such occasions of card-coincidences and not at other times. It is suggested that he can act only in reinforcement. More study of this phenomenon too is desirable.

SUMMARY

The average of 5.48 obtained from 1,224 runs of GESP and DT was highly extra-chance, with a critical ratio of 8.37. The conditions excluded possibility of sensory cues, and double checking and rechecking made errors improbable.

A previously "successful" and an "unsuccessful" experimenter were compared in these tests, both taking equal part. The averages in the GESP work obtained by the "successful" experimenter were close to those obtained by him in former experiments. The results obtained by the "unsuccessful" experimenter were only slightly higher than those obtained by him in his individual work. In view of the fact that the conditions and the subjects were the same for both experimenters, one conclusion that is evident is that the mere desire to succeed as an ESP experimenter is not enough. It is not possible to state at present upon what individual differences in the experimenters the discrimination of the subjects is based. The research shows only that if differences in personality were responsible for the results, they were at least not dependent on direct contact with subjects during the tests; experimenters and subjects had no such personal contact.

In this case the "card by card" GESP technique proved superior to the "whole deck" DT technique, but it is not possible to deny that telepathy may have played a part also.

The phenomenon of the U-curve in DT work did not occur under the conditions followed, and it is suggested that the difference in these conditions may be responsible for its absence.

1,527 such coincidences, as against an expected 1,530. Likewise, the already-mentioned check for equal numbers of symbols in the run gives further reassurance on the point in question.

THE EFFECT OF RATE OF MOVEMENT IN CARD MATCHING TESTS OF EXTRA-SENSORY PERCEPTION*

CHARLES E. STUART

ABSTRACT: After their normal tapping rate was ascertained, subjects were instructed to match correctly unseen ESP symbols in time to the beat of a metronome. Two techniques, blind matching and screened touch matching, were used with 41 random subjects. When the metronome rate approximated the normal tempo, the subjects' scores averaged 5.67 per run. When the matching was done at rates other than the normal the average successful hits per run was 4.96. The difference noted is highly significant, and due mainly to significant extra-chance scoring at the tempo rate. It is suggested that the difference in scoring level noted is due to distraction of attention from the primary task of matching when imposed non-tempo rates require effort to keep in time.

Mr. Stuart is a member of the staff of the Parapsychology Laboratory at Duke University.

ONE OF THE difficulties in the path of study of conditions which affect extra-sensory perception in a reliable manner is the fact that the most necessary conditions seem to be subjective states difficult or impossible to control directly in the laboratory. Rhine has observed that "good *abstraction* is required for success in most percipients, along with effort and attention to the task" (2, p. 164). Abstraction may be aided by a comfortable laboratory atmosphere and social situation. However, the direction of effort and attention to the task not only must be left entirely to the subject, but the experimenter is unable even to describe the task explicitly. In the verbal calling of the top card of a deck, the subject may direct attention to auditory imagery, visual impressions occurring in the gray field of his closed eyes, or slight images occurring as he gazes at a blank wall. When he is matching one card to another he may attend to similar visual imagery or to kinaesthetic experiences in the hand and arm doing the matching. But whether attention so directed is effective toward, or merely distracting to, successful per-

* Presented at the thirty-third annual meeting of the Southern Society for Philosophy and Psychology, April 15-16, 1938, at Knoxville, Tenn.

formance is not a question that may be settled upon the evidence now available.

When matching unseen symbols on cards subjects occasionally adopt spontaneously a rhythmic performance. Such an observed change obviously indicates a shift of attention to that particular mechanical feature of the task. The observation suggests that a rhythmic rate of movement might be used as an objectively measurable method for shifting the attention of the subject experimentally, and thus enabling the effect of such a shift to be observed.

Stern introduced the concept of *personal tempo* in 1900, (4) and has more recently remarked "Each individual has his *personal tempo,* i.e., an optimal time of rhythmic events in life and experience. The less the speed of activity is interfered with by external causes, the greater is the sway of this natural rhythm." (3) Allport and Vernon (1) found that, while this generality does not hold true quantitatively for all individual events, rhythmic motor activities are highly intercorrelated within the individual. These observations would suggest again that attention to a given motor activity might be experimentally varied by imposing different rhythmic rates of movement, some of which might be optimal for that task and others which might so interfere with the normal speed of activity as to constitute an effective distraction.

The following study is an attempt to apply these suggestions to ESP tests wherein the subject's response is a motor act. By imposing varying rates of movement some difference should be observed between results at the subject's spontaneously preferred rate and other rates not so preferred.

SERIES I—BLIND MATCHING PROCEDURE

The subject, knowing no more than that he was about to be tested for extra-sensory perception, was shown the symbols of a deck of ESP cards if he was unfamiliar with them. He was then given a pencil and asked to tap out a three-quarter-time rhythm on the table top. If he did not understand, or had difficulty in achieving a regular rate, the experimenter illustrated the pattern of the time, but avoided a regular rhythm. After the subject had tapped fairly regularly for thirty seconds, his rate was timed for thirty seconds. This rate was recorded as his "normal tempo." Five key cards were then placed before him. These were five ESP cards with the customary five symbols, pasted face down in a row on a strip of heavy paper. Their order was known to the experimenter, but the cards were never exposed by the experimenter and pains were taken to avoid any indication of the order.

The subject was handed a shuffled deck of ESP cards (25 cards consisting of 5 cards of the same symbols as the key cards) and was instructed to match them face down to the key cards. The matching was to be done smoothly and rhythmically in time with the beat of a metronome. The suggestion was made that the whole arm and shoulder be used in the matching movement. A card was to be matched at every second sound of the metronome. That is, for the subject to match cards at a rate of sixty per minute the metronome was set at 120. (A *beat*, hereafter, will mean two sounds of the metronome.)

The metronome was first set beating at a rate twenty beats slower than the subject's normal tempo, then at the normal rate, then at a rate about twenty beats faster. This pattern was repeated a number of times, with various other rates occasionally interposed. The subject matched twenty-five cards at each rate. The experimenter checked the cards and recorded the number of hits after each run, and reshuffled the cards for the next run. The subject was told the number of hits made after each run. At no time, other than the few cases in which he was introduced to the symbols at the beginning of his work, did the subject see the face of a card during the experiment.

Twenty-four subjects were used. The results of two of these (the experimenter and a colleague) although consistent with the general trend, are omitted from this report because the results might be affected by knowledge of the nature of the experiment. The remaining twenty-two are randomly chosen subjects varying in age from 20 to 40. All knew that the experiment was a test of extra-sensory perception, and assumed, without being explicitly instructed, that they were to make as high a score as possible on each run. All knew that a score of five was a chance score. The experimenter tried to maintain a consistent attitude toward all runs, encouraging the subject if his score was low, congratulating him if the score was high. The social situation was made as light as possible. If the subject tended to be too seriously involved in his task, the experimenter engaged him in light conversation or repartee.

No definite procedure or number of runs was prescribed for each subject except that the tempo rate was embedded in a series from below tempo to above tempo. The experiment was discontinued upon signs of fatigue or boredom on the part of the subject. It was limited in some cases simply upon his report when asked whether he wished to continue.

SERIES II—PROCEDURE WITH SCREENED TOUCH MATCHING

The second experimental series was similar in procedure to the first, but utilized the screened touch matching (STM) technique instead of blind matching for the subject's response. In this technique the subject is seated at a table opposite the experimenter. Between them is an upright wooden screen large enough to block the subject's view of the experimenter and the area of the table immediately before him. Between the bottom of the screen and the table top is an aperture about two inches high. Within this aperture are five shallow boxes, each just large enough to hold a card. A smaller upright screen, a few inches back of the aperture on the experimenter's side, completely blocks vision from the subject's side through the opening but allows the experimenter to look down at the shallow boxes.

The key cards were placed face up in the shallow boxes. The subject was instructed to tap the face of a card with a pointer in time with the metronome. The experimenter, holding the pack of cards in his left hand, dealt out each card in line (behind the screen) with the box indicated. In order not to break the subject's rhythm, the cards were dealt so that any sound they might make coincided with the second sound of the metronome. The experimenter recorded the hits, reshuffled the cards, and changed the metronome, all behind the screen. In this experiment the subject had the five symbols presented visually before him, but at no time did he see the cards being used behind the screen.

A definite patterning of the presentation of metronome rates was adhered to somewhat more rigidly than in the former experiment. The general pattern was one run each at twenty beats below normal tempo, ten beats below normal tempo, the normal tempo, the normal tempo repeated, ten beats above, and twenty beats above. Repetitions were presented in the same order. The main variation in this was the presentation of fifteen beats above and below, instead of twenty.

Ten subjects were used for Series II. Two of these were subjects used in the previous experiment.

SERIES III

The third series was conducted in the same manner as the second, with the exception that the rates were presented in a decreasing as well as in an increasing series. Ten subjects were used, one of whom had been used in Series II.

The order of the three series is chronological—that is, Series I was completed before Series II and Series II before Series III.

RESULTS

Scores made by the subjects at their individual preferred rate of movement, as measured at the beginning of each experimental period, were pooled as runs at the "normal tempo." Runs made at rates less than five beats per minute slower than, or more than five beats per minute faster than, the normal tempo were pooled to give the "non-tempo" scores.

In Series I with the blind matching technique 141 runs at the normal tempo rate yielded 799 hits, an average of 5.67 per run. At rates other than the normal tempo 278 runs yielded 1,377 hits, an average of 4.95 per run. The difference between these averages, .713, divided by the standard deviation of the difference, .227 gives a critical ratio of 3.23, which indicates odds of about 2,600 to 1 that some factor other than chance variation must account for the difference. An evaluation of the averages in terms of an expected average of 5.00 per run gives a critical ratio of 3.96 for the tempo group, 0.57 for the non-tempo group. The above difference is mainly due, therefore, to the significantly extra-chance scores at the tempo rate.

In Series II with the screened matching technique 140 runs at the normal tempo rate yielded 774 hits, an average of 5.53 per run. At rates other than the normal tempo 227 runs yielded 1,099 hits, an average of 4.84 per run. The difference between these averages, .687, is 3.13 times its standard deviation, .219, which indicates the averages to be significantly different. Evaluation of the averages in terms of their deviation from an expected average of 5.00 per run gives a critical ratio of 3.13 for the tempo group, and 1.19 for the non-tempo group. The difference is again indicated to be due mainly to the significantly extra-chance scoring at the tempo rate.

In Series III with the screened matching technique, 82 runs at the normal tempo rate yielded 468 hits, an average of 5.93 per run. At rates other than the normal tempo 130 runs yielded 662 hits, an average of 5.09 per run. The difference is significant, and is due to significantly extra-chance scoring at the tempo rate.

The total scores at the tempo rate yielded an average of $5.69 \pm .116$. The total scores at non-tempo rates gave an average of $4.96 \pm .082$.

Table I gives the individual subject scores for the three series. The first column, marked "Tempo" is the tapping rate of the subject ascertained before the matching tests. It will be seen that subjects used for more than one experimental period showed considerable variation in preferred rate. Subject V, for example, was timed at 60, 64, 60, 50, and 69 in five sessions. No correlation was noted between preferred

TABLE I

SCORES OF INDIVIDUAL SUBJECTS

Series I: Blind Matching

Subject	Tempo	Tempo			Non-Tempo		
		Runs	Hits	Ave.	Runs	Hits	Ave.
A	40	3	27	9.0	6	24	4.0
B	68	2	18	9.0	3	15	5.0
C	50	2	16	8.0	9	51	5.7
D	60	3	22	7.3	9	51	5.7
E	50	4	27	6.8	12	51	4.3
F	50	6	40	6.7	9	43	4.8
G	60	11	74	6.7	5	25	5.0
H	66	4	26	6.5	11	53	4.8
I	40	5	32	6.4	15	57	3.8
J	70	3	19	6.3	18	106	5.9
K	60	6	36	6.0	17	97	5.7
L	55	5	30	6.0	12	70	5.8
M	60	6	36	6.0	16	73	4.6
N	60	12	65	5.4	29	133	4.7
O	50	5	25	5.0	9	37	4.1
P	70	11	54	4.9	13	61	4.7
Q	56	17	84	4.9	25	126	5.1
R	76	8	39	4.8	5	16	3.2
S	55	4	19	4.8	8	43	5.4
T	60	13	63	4.8	29	149	5.1
U	60	6	26	4.3	8	53	6.6
V	60	5	21	4.2	10	42	4.2
Totals		141	799	5.72	278	1377	4.95

Series II: Screen and Touch Matching

Subject	Tempo	Tempo			Non-Tempo		
		Runs	Hits	Ave.	Runs	Hits	Ave.
W	64	5	31	6.2	17	81	4.8
X	60	6	37	6.2	12	54	4.5
Y	60	6	36	6.0	10	54	5.4
Z	56	9	54	6.0	16	91	5.7
Z	60	6	44	7.3	10	40	4.0
Z	48	6	31	5.2	8	45	5.6
Z	62	6	32	5.3	10	45	4.5
AA	46	6	34	5.7	12	64	5.3
BB	88	14	79	5.6	19	95	5.0
CC	64	7	39	5.6	12	67	5.7
V	64	9	50	5.6	16	64	4.0
V	60	6	33	5.5	10	42	4.2
V	50	8	34	4.3	12	52	4.3
V	69	8	46	5.8	8	43	5.4
DD	58	6	32	5.3	12	61	5.1
DD	66	6	32	5.3	12	55	4.6
A	40	11	57	5.2	13	61	4.7
A	50	6	27	4.5	10	48	4.8
EE	60	9	46	5.1	8	37	4.6
Totals		140	774	5.53	227	1099	4.84

Series III:

Subject	Tempo	Tempo I. S.			Tempo D. S.			Non-Tempo		
		Runs	Hits	Ave.	Runs	Hits	Ave.	Runs	Hits	Ave.
Y	60	4	18	4.5	4	19	4.8	12	56	4.6
Y	50	6	42	7.0	6	35	5.9	18	104	5.8
FF	64	8	45	5.6	8	50	6.3	22	95	4.3
GG	50	6	38	6.3	4	24	6.0	16	74	4.6
HH	50	2	10	5.0	2	16	8.0	8	52	6.5
II	50	2	11	5.5	2	15	7.5	8	36	4.5
JJ	55	2	9	4.5	2	14	7.0	8	47	5.9
KK	40	4	17	4.3	4	29	7.3	13	80	6.2
LL	58	2	9	4.5	2	15	7.5	6	23	3.8
MM	55	2	10	5.0	2	14	7.0	6	32	5.3
NN	46	2	13	6.5	2	10	5.0	7	32	4.6
NN	56	2	15	7.5	2	8	4.0	6	31	5.2
Totals		42	237	5.64	40	249	6.22	130	662	5.09

tempo and scoring at that rate. The very unreliable individual averages (owing to the small number of runs at the tempo rate) would make the bearing of any correlation questionable.

In Series III the tempo rate was imbedded in both an increasing and decreasing rate series of runs. The higher scoring for the latter is striking, although not significantly different from the average of the former.

Although no attempt was made to get significant scoring from individual subjects in this study, the record of a subject whose previous scoring indicated consistent ability in ESP is of illustrative interest. Table II gives the scores made by Mr. J. D. MacFarland in Series III.

TABLE II

EXAMPLE OF PERFORMANCE BY AN "ESP SUBJECT"

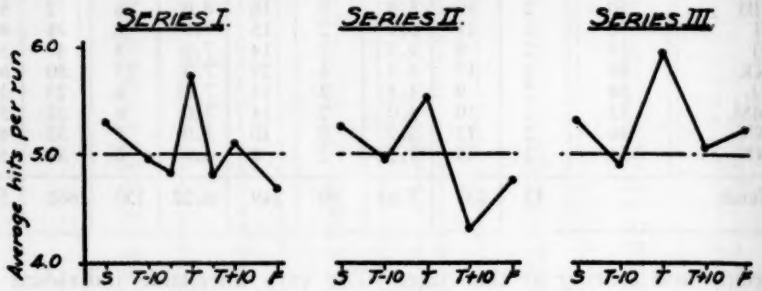
Subject: J. D. MacFarland

Tempo Rate: 64 Sept. 21, 1937

Metronome Setting	Increasing Rate	Decreasing Rate	I. R.	D. R.	I. R.	D. R.
98	7	4	*	3	*	*
108	3	2	5	4	2	3
128	7	6	7	11	5	6
128	5	4	8	7	2	5
148	7	4	3	3	5	4
158	7	*	5	*	5	*

Average at Tempo Rate: 6.08.
Average at Non-Tempo Rates: 4.22.

When the non-tempo rates are analyzed in groups about the tempo rate no striking consistency is noticeable. Table III and Graph 1 present the results in comparison with the metronome rate. A "slow" rate (S on the graph) is any rate less than 10 beats below the tempo rate. A "fast" rate is any rate more than 10 beats above the tempo rate. In Series I observations were made at five beats below and above tempo; in Series II and III no trials were made at these rates.



GRAPH 1

TABLE III

COMPARISON OF AVERAGE SCORE WITH METRONOME RATE

		Slow	T−10	T−5	T	T +5	T +10	Fast
Series I..	R.	36	82	34	141	37	53	36
	H.	191	407	164	799	178	269	168
	Av.	5.31	4.96	4.82	5.72	4.81	5.08	4.67
Series II..	R.	63	56	....	140	....	50	58
	H.	331	277	....	774	....	215	276
	Av.	5.25	4.95	....	5.53	....	4.30	4.76
Series III..	R.	27	39	....	82	....	38	26
	H.	143	191	....	486	....	192	136
	Av.	5.30	4.90	....	5.93	....	5.05	5.23

None of the averages other than those at the tempo rate (all of which are significant) meet a criterion of a 2.7 critical ratio. The low point, T + 10, in Series II, reaches a critical ratio of 2.5, but this low point does not recur in Series III.

DISCUSSION

The first question arising is that of the occurrence of extra-sensory perception. Have these tests shown evidence of extra-chance scoring under conditions ruling out sensory or rational bases for response?

This question may be approached without reference to the working assumptions of the experiment. The total 998 runs under all conditions gives an average of 5.2, which is reliably above the expected 5.0. The experiment yielded extra-chance scoring. May hypotheses other than extra-sensory perception explain the results? Most of these can be tested directly.

1. Is it possible that in Series I, when the subject handled the cards, and wherein the absence of cues as to the order of the key cards depended upon the experimenter's vigilance, that an occasional sensory or rational cue was available, and thus provided the positive deviation? If so, we should expect most of the deviation to result from that experiment and little or none from the screened matching experiments in which these possibilities were wholly ruled out.

The data show, however, that the same averages resulted from both techniques, and that Series II and III are independently significant.

It might be observed further that the different scoring level at tempo and non-tempo rates would suggest, on a sensory-cues hypothesis, that available cues were used only at the tempo rate and ignored otherwise; a highly unlikely possibility.

2. Is it possible that counting errors on the part of the experimenter accounts for the positive deviation?

The method of recording in the screened matching procedure was as follows: After the subject had finished the run there were, on the table before the experimenter, five piles of cards, each in front of the symbol the cards were intended by the subject to match. Each of these piles was then turned face up and separated into two groups, the cards which matched the key card in front of the pile, and those which did not. As the correct matchings were thus sorted out they were counted. When the sorting was finished the correct matchings lay face up directly adjacent to the key symbols, and the first counting could be checked at a glance. Occasional carelessness might result in the oversight of a correct card in the sorting of the piles, but error in the counting of the correct cards after sorting is extremely unlikely.

Furthermore, if the error hypothesis be tested rigorously against the data it will be noted that the average of 5.7 at the tempo rate must be explained as well as the chance averages at non-tempo rates. This would mean that, while errors at the non-tempo rates were either negligible or compensated, at the tempo rate, in order to produce the average observed, two out of every three runs would have to be assigned through error one more correct matching than actually occurred. The writer can formulate no reasonable hypothesis which would explain such

a sharp fluctuation from negligible error to what would, as an error, constitute sheer hallucination.

3. Another alternative hypothesis may be stated, test of which is not directly available in the results. In the STM procedure in Series III printed ESP cards were used. These cards are imperfect in many ways, and sensory cues from the backs are available under certain lighting conditions. The screen makes these possible cues unavailable to the subject. And so long as the experimenter pursues his mechanical function of placing each card before the symbol designated by the subject's pointer, the state of the cards are of no concern to him. But since the experimenter is naturally interested in some consistent outcome of his research, and may favor at some time or other an hypothesis which would be verified by differential scoring, it can be proposed that the experimenter may occasionally react to an unconscious sensory cue, and place the card, not where the subject points, but before the correct symbol. Since in these experiments there were no objective records of the order of the subject's responses direct check of such a motivational error can not be effected from the data.

In answer to such an hypothesis it is possible to offer the general argument that an experimenter with psychological training usually is familiar enough with motivational error to strive consciously for exactness in spite of it. The ESP experimenter, specifically, learns to avoid competing with the subject in identifying the cards. His attention is, moreover, completely occupied with the moving pointer below the screen. Vision of the cards in his hand is necessarily peripheral, and no student of ESP card imperfections has yet reported identifying them with peripheral vision. There is, fortunately, one bit of evidence available which bears upon the suggestion of motivational error. When Series III was projected, the writer was strongly convinced, on what seemed excellent theoretical grounds, that the tempo rate when approached from a faster rate would give chance results. The results persistently denied the expectation.

With sensory and rational cues ruled out, and experimental error extremely unlikely, the hypothesis of extra-sensory perception appears accordingly necessary to explain the results.

In reports heretofore of extra-sensory perception, attention has been frequently directed to extra-chance scoring *by a particular subject.* Individual subjects capable of consistent performance are selected not only to demonstrate that the capacity is that of a human individual, but also to provide a convenient localization of the ability under study. Here no attempt was made to select subjects. Rather the "selection"

was that of an objectively describable experimental situation. It was hoped that success with random subjects would make the tempo rate matching a useful contribution to ESP test methods. That is, matching at the tempo rate might provide such an optimal condition for ESP that significant scores could be reliably expected from any group of subjects.

This expectation was verified in the above results. The maximum and significant level of scoring for all three series occurred at the tempo rate. And no other hypothesis than chance is necessary to account for scores at the non-tempo rates.

As has been noted, in Series I the only fixed order of presentation of rates was that the tempo rate should be imbedded in a series from a slower to a faster rate. The experimenter felt free to vary the order in Series II as well, although patterns were more frequently adhered to in this series. The question arose whether the results might be due not to any variation of rate as such but by the experimenter's recognition of slight changes in the social situation, in the subject's expressions of confidence, etc., that would allow him to present the tempo rate under these intuitively judged better conditions. Series III, however, suggests that this hypothesis does not hold. Patterns were adhered to exactly, and the averages, 5.95 at the tempo rate, and 5.04 at non-tempo rates, were strictly comparable to the previous series in which freedom of pattern of presentation was admitted.

In view of recent demonstrations of the importance of the role of the experimenter in ESP tests, the question might be raised whether the experimenter provided by some differential expression of attitude toward tempo and non-tempo runs, a favorable and unfavorable situation in line with his expectations. Again, the results of the decreasing rates in Series III, in spite of the expectations of the experimenter, would seem to bear against such a proposal.

The question remains to be considered whether the tempo rate actually produces or increases extra-chance scoring at that rate, or whether these averages represent the general extra-chance scoring level of which the group is capable, but which is inhibited by the non-tempo rates. That is, what average would be expected from these random subjects under a similar good ESP test without the imposed rate? Since not all the subjects were given comparable tests without the imposed rate, nor when such tests were given were the conditions objectively classifiable as "good ESP conditions," this question cannot be answered directly. But relevant bearing is suggested by the following study.

From exploratory tests given by this writer to subjects never before tested, the first four runs made by 25 randomly selected subjects were taken as a sample. The subjects were mainly individuals visiting the laboratory who were interested in seeing how the tests were conducted. Such conditions are generally highly favorable. The average score found was 5.57.

If this observation affords a valid standard of comparison it suggests strongly that the tempo-non-tempo difference is due to inhibited scoring at the non-tempo rates.

But what is the nature of this inhibition? Spontaneous observations by the subjects often take the form "That's too fast," or "That's too slow." The word "too" suggests that the subject is thinking of his matching task and the imposed rhythm as separate parts of the situation, and that the rate is distracting to the matching task which the subject has formulated. The occasional spontaneous "That's better" when the normal tempo is presented also suggests that there was something disturbing about the previous rate.

Since many observers have noted that distractions lower ESP scores, while apparent freedom from distraction enhances the subject's scores, it is a simple speculative step from the above observations to the working hypothesis of the experiment, that the presentation of the various rates changes the amount of attention the subject may give to the primary task of "hitting" the unseen ESP symbol. At the normal tempo little attention need be given to keeping in time with the imposed rate, and full attention may be given to the task of matching symbols correctly. At other rates it appears that more effort must be directed to keeping in time with the imposed rate, which condition effectively distracts attention from the test of matching the cards correctly.

SUMMARY

Blind matching and screened touch matching ESP test techniques were used with 41 subjects. The normal preferred tapping rate of the subjects was ascertained. They were then instructed to match ESP symbols in time to the beat of a metronome. The metronome was set at various rates ranging about the normal tempo of the subject.

When the metronome rate approximated the normal tempo, a significant average of 5.67 was observed. When the matching was done at rates other than the normal the average successful hits per run was 4.96, scarcely different from chance scoring.

It is suggested that the difference in scoring level noted is the result of inhibition of scoring at the non-tempo rates. The inhibition is effected

by distraction of attention from the test of matching cards correctly to the secondary test of keeping in time with the metronome.

REFERENCES

1. ALLPORT, G. W., AND VERNON, P. E. Studies in expressive movement. New York: Macmillan Co., 1933.
2. RHINE, J. B. Extra-sensory perception. Boston: Bruce-Humphries, Inc., 1934.
3. STERN, W. General psychology. New York: Macmillan Co., 1938.
4. STERN, W. Psychologie der individuellen Differenzen. Leipzig: 1900.

AN UNUSUAL CASE OF EXTRA-SENSORY PERCEPTION*

RALEIGH M. DRAKE

ABSTRACT: An eleven-year-old boy, physically retarded and with an I.Q. of 55, was found to repeat spontaneously words and numbers thought of by his mother but not overtly expressed. The boy responded with unusual accuracy when verbally urged by the mother. ESP cards were utilized with similar results. The words used by the mother to attract the child's attention were then limited in number, but high scoring continued (upon one occasion an average of 21 correct hits per run for 14 runs was observed). Finally the child was induced to respond with a significant average of 7.6 per run to a mechanical ready signal operated by the experimenter without verbal urging by the mother. The accuracy of response to both verbal and mechanical signals declined rapidly toward the end of the study.

Dr. Drake is Professor of Psychology at Wesleyan College in Macon, Ga.

IN THE FALL of 1937, I received a letter from a former student who asked if I might be able to help a boy who was having unusual difficulties in school, especially in reading. It was also timidly mentioned, no doubt because the student knew of my skepticism regarding the topic, that the boy possessed the "ability to read minds, having on numerous occasions been able to tell other people exactly what they are thinking, or what they saw the day before." I expressed my willingness to examine the boy both as to the cause of the reading disability and the cause of the supposed mind-reading ability, which I fully expected to be due to some minimal sensory cue. The case turned out to be unusual in many respects especially in its bearing on the hypothesis of extra-sensory perception.

DESCRIPTION OF THE SUBJECT

The subject (whose nickname "Bo" I shall use for proper designation) is an eleven year old boy who, at the time of birth, received a cerebral injury which has caused marked physical and mental retardation. By the Stanford Revision (1915) his I.Q. is 55; by the Pintner-Cun-

* Presented at the thirty-third annual meeting of the Southern Society for Philosophy and Psychology, April 15-16, 1938, at Knoxville, Tenn.

ningham Primary test it is 53. He is unable to do second-grade school work because of his distractibility and general inability to perceive relationships and remember facts. There is no evidence of unusual, or even average, memory for any specific kind of material which is sometimes noted in retarded cases, for an analysis of the memory items in the Stanford-Binet test shows a failure to repeat one of the three sentences at age VI, or one of the five digit numbers at age VII.

Muscular control and co-ordination are poor as evidenced in the halting manner of walking down steps, by inability to run, roller skate or learn to ride a bicycle, and by irrelevant nervous muscular hyperactivity. No special artistic, mechanical, musical, mathematical, memorial, special, or other abilities have been shown in any way which would account for the phenomenon under investigation, or form the basis for an educational program.

Because the boy made no progress in school the mother coached him in his school work. It was thus that she discovered that he would spontaneously tell her words or numbers which she had not overtly expressed. For a while he was thought to be a lightning calculator because no matter what the row of figures given he would immediately give the answer, provided it was in the mother's mind, but he could do absolutely nothing if left alone. As reported by the former student, "The child cannot read unless someone is at his side looking on his book. He reads well then, but when left alone, he cannot read."

Physically the child is underweight, approximately twenty pounds, and is correspondingly short in stature. The available energy from the large amount of food eaten is not properly assimilated, which has led to the administration of thyroxin by the physician under whose care he has been since October 1937. Over a period of about eight months this treatment has produced no noticeable improvement in any respect, the weight being almost exactly what it was eight months previously.

The personality of the child might be characterized by irrelevant, distractible, and hyperactive extroversion. The attention wanders from one thing to another with no logical connection between topics while some topics are returned to repetitiously. Always there is over-reaction in the form of nervous muscular activity, gestures, verbal emphasis, and verbigeration. There is little anticipation of future rewards on the one hand or little memory for past joys or sorrows on the other. On one occasion when he was to leave home for a few days to stay in my home he showed no emotion at leaving his parents or any homesickness or anxiety to return. Even at the moment of seeing his parents after an absence of two days the meeting was no more than casual on his part.

He is an active observer of the activities of other boys, usually younger than he, but he makes little effort to participate co-operatively in their games. He appears to be interested in everything that goes on and makes comments which are more or less relevant. His comments are, however, frequently of the perseverative type with verbal repetition and fixation of ideas constantly recurring.

EXPLORATORY INVESTIGATION

The first time I saw the boy and his mother (October 9, 1937), the latter always acting as agent or sender, they tried to transfer words, page numbers and symbols, but never with much success. The mother would look at a word, sentence, or number and ask the child to tell us what she was thinking of, the experimenter always knowing what the mother was reading at the time in order to check against the call. There were no spontaneous *whole* answers. Instead there would be a lot of guessing at words or attempts to spell them out, with the mother continually stimulating him with "what's the next letter?", "what's this word?", or "what's the next number?" frequently adding his nickname "Bo" as an added stimulus. Occasionally he would give surprisingly accurate responses with, or without, specific verbal stimulation. On one occasion the mother read from a book which I selected as entirely new to her (Terman's Condensed Manual) and the boy repeated a large majority of the words, all of the time paying no attention to his mother while I carefully watched her lips for incipient movements and listened for minimal verbal cues, none of which were detected. Dr. Bruce of the mathematics department at Wesleyan College witnessed this demonstration and was much impressed with the boy's ability.

The second time I saw the child (October 12, 1937) most of the time was given to administering the Stanford Revision and only a few casual attempts were made at investigating his unusual perception. I did, however, write down twelve numbers, which were seen only by the mother, and the boy was asked one at a time to give the number on which the mother was concentrating. This he did with only fifteen attempts, or with only three errors. At no time could I hear any auditory cues from the mother and when I held a piece of paper directly in front of his eyes, so he could not see the mother at all, he seemed to respond even quicker. On no occasion during all of the experimental tests have I seen him make any attempt to watch the mother for possible signals. In fact he is so busy most of the time with irrelevant things that he scarcely glances at his mother. He is so distractible that it is almost impossible to keep his attention on anything for more than

a moment or two at a time without having to almost forcibly bring it back to the task in hand.

On another occasion (January 5, 1938) I wrote six groups of three numbers each for the mother to look at and for the boy to call without in any way seeing them. The numbers that I wrote down and the exact calls made are given below.

TABLE I

NUMBERS AS STIMULI

Numbers on which mother concentrated	Numbers called by the subject	Number of trials to get each correct
8	1, 3, 2, *2*, 8	5
6	9, 10, 3, *9*, 12, 8, *9*, *9*, *3*, 6	10
7	3, 2, 9, *3*, *9*, *2*, 8, 6, 4, 5, *2*, *7*	12
9	9	1
3	2, 3	2
4	4	1
10	5, 10	2
11	9, 3, 4, 8, *9*, 11	6
3	12, 10, 9, *12*, 2, 4, 13, 1, *9*, *13*, *1*, 3	12
8	4, 10, 9, 6, 8	5
12	9, 2, 3, 5, 12	5
3	4, 6, 3,	3
5	1, 0, 2, 9, *2*, 4, 5	7
6	6	1
7	7	1
10	10	1
5	11, 12, *12*, 14, 9, 4, 5	7
3	3	1
Total 18		Total 82 Minus repetition 16 Net total 66

It is difficult to evaluate these results because we do not know what the chance expectancy is for random numbers of this description. But it is rather striking that 6 out of 18 numbers were correct the first call. It will be noted that the subject was free to call until he had guessed correctly. Deducting repetitions, italicized in the table, all eighteen numbers were called in 66 trials.

One interesting observation of the effect of punishment may be noted. At the thirteenth number, which was 5, the mother insisted that the boy was not trying and had promised him, in some of the earlier trials, that he would be switched if he did not do better. After 4 was

called, which was wrong, the mother got a small stick and switched him enough to frighten him, after which he immediately gave the correct answer of 5. The next three consecutive calls were also correct making, in all, four in quick succession which was considerably better than any run heretofore. This tendency to do well for short spurts of time when forced to pay attention to the task in hand was subsequently noticed many times, but is never more than momentary, for concentration is not easily achieved by the child.

At this point it was felt that some more systematic investigation should be undertaken and methods introduced which would measure the amount of success and adequately allow for the chance or guessing factor. For these reasons the Duke ESP cards were introduced and employed in practically all of the experiments which followed. Procedures and results are reported in chronological order to show the progress that was made and the difficulties which arose.

The ESP cards were first used on January 5, 1938. A large amount of the experimental period of about an hour was consumed with teaching the boy the names of the cards and trying different methods. For instance, we first tried to get him to call the cards, with the mother acting as agent, before he knew what symbols were on the cards, i.e., before he had seen an ESP deck of cards. The first card pulled by the mother, and seen by both the mother and experimenter, was a star which the boy immediately called a "star." The second card pulled was a cross which he could not call after much hesitation and repeating "what card mother"?

Finally the mother asked him to spell the word which was done in the following way. The first letter he gave a "c" after which the mother said "c what?" and he replied "cr." The mother then said "cr what?" and he responded with "o." The mother questioned him further with "cro what?" and he responded with five errors in succession of "r," "w," "n," "a," and "e." He got the first "s" but was never able to give the last "s" because the mother gave up the attempt at this point. Altogether twelve cards were pulled but no further correct whole answers were given and the attempts to spell out the names of the symbols on the cards were very unsuccessful.

Thinking that possibly it made some difference in the way which the mother concentrated, as for instance whether she visualized or concentrated on the name of the symbol or the geometrical shape of the symbol, we asked the boy to draw the symbol. All such attempts were, however, total failures perhaps partly, as we found out later, because he was able only with difficulty even to copy a square, a cross, or a star.

We then took a second ESP deck and laid out the five different cards on the table in front of him and asked him to give his response by pointing to a card before him on the table. Most of these attempts were also failures.

Finally we tried the conventional procedure of the agent concentrating on a card and the percipient making a call. A few successes followed with this method but with all the coaxing, urging, and even threatening by the mother we succeeded in getting, by all methods, only 53 total responses, thirteen of which were correct, or about 6 out of 25 right. During this time the boy was informed whether or not his call was correct and the mother usually gave him two or more trials but no record was kept of any but the first call given.

On January 30, 1938, we managed with great difficulty to get 90 calls in between his crying, fussing, rebelling, fidgeting, and counter talk; 66 of these were correct, making an average of about 18.3 correct per run of 25. This, by the way, was the first time any tests were made in his home, all others having been made in my office or laboratory.

At this point the child spent two days in my home during which time numerous efforts were made to get his co-operation with various adults and children acting as agents. With myself as agent in four trials scores of 4, 9, 6, and 4 were obtained. With Mrs. Drake acting as agent in eight trials the scores of 7, 2, 6, 4, 11, 5, 4, and 4 were obtained, or an average for the series of 5.4. My five-year-old son was sender for one run with which a score of 5 was secured. My ten-year-old daughter tried to act as sender for one run but could not get enough of the subject's attention to finish the 25 cards in the deck. During most of these trials the boy ignored us, or told us to "shut up" even when nothing was said, or rode about the room on a small tricycle, or shot down imaginary robbers. He tended to adopt a method of least resistance by calling mostly waves and "boxes," the latter being his designation of a square.

On February 1, 1938, he was taken back to his home and, with the mother acting as agent, we managed with bribery, threats, and catching him in the right mood, to get 595 calls with 373 correct, which would amount to about 16 correct per run of 25. Although discouraged from so doing the mother persisted in giving second and third chances when the first call was incorrect and to put all miscalled cards in a separate pile from which she would draw when the main deck was exhausted. This resulted in securing 25 regular calls plus the number of incorrect calls for every run through the deck, or from 25 to 35 calls per run. These are listed in Table II.

TABLE II

ESP CARDS AS STIMULI

Date	Run	Number of cards pulled	Correct calls, 1st chance	Number of trials, 2nd chance	Number correct, 2nd chance	Number of trials, 3rd chance	Number correct, 3rd chance
Jan. 30	1	29	20	no record	..	no record	..
"	2	29	22	6	4	..	..
"	3	32	24	6	2	3	2
Feb. 1	1	35	16	10	10	..	..
"	2	26	15	9	5	4	4
"	3	29	20	6	5	1	1
"	4	27	19	6	3	3	3
"	5	30	20	7	7	..	..
"	6	27	21	2	2	..	..
"	7	35	23	5	5	..	..
"	8	26	15	8	6	2	2
"	9	26	15	4	4	..	..
"	10	39	28	4	4	..	..
"	11	30	20	8	8	..	..
"	12	31	18	5	5	..	..
"	13	31	17	6	4	1	..
"	14	26	17	7	6	1	1
"	15	26	12	9	9	..	..
"	16	40	28	7	6	1	..
"	17	12	8	..	..	..	..
"	18	36	25	5	5	..	..
"	19	25	17	6	5	1	1
"	20	31	19	7	6	1	..
Totals		678	439	127	111	18	14

From this table it will be seen that out of 678 trials, 439 were correct, or on the average 64 per cent of the first calls were correct. For second chances 87 per cent were correct and for third chances 78 per cent were correct. Records for second and third chances are given simply beause that is the way the experiment proceeded, over the protests of the experimenter, and no significance is attached to these results although taken at face value they are in themselves impressive. Dr. Bruce was a witness to all of the trials of February 1 and was positive there were no verbal cues responsible for the phenomenon. Most of the trials were conducted with no special precautions to isolate the boy and the mother but in a few trials he was blindfolded, during which time he did better than when unblindfolded, and for one of the best runs he was in another room with the door open just wide enough to allow an observer to see the cards which were pulled by the mother. At various times Dr. Bruce and I would be between mother and son but neither of us heard any auditory cues other than the usual verbal

signal, to indicate readiness for the next call, of "what is the next card?", "come on let's go," "Bo," "go on," or "and."

For one of the runs she used nothing but the word "and" because the subject was calling so rapidly and working so well at the moment that saying only the one short word saved time. It seemed that the faster he could go and the less said the better he worked and the higher his score. This same tendency was noticed in all of the later tests; he would have spells during which he would go extremely fast and accurately for a short time and then relapse into refusal to call or make numerous mistakes.

EXPERIMENTAL

Up to this time the control of the sender was not very adequate. The mother would say anything that came into her mind to stimulate the boy to make a call. It could readily be asked whether this might be a cue to the five symbols on the ESP cards. Conceivably five different verbal cues could be learned by an alert recipient to correspond to the star, square, circle, cross, and waves on the cards, although such cues would have to be very simple for this child to learn them. However, in order to eliminate, as far as possible, all sensory cues we attempted to apply further precautions. In most of the series of runs beginning with February 13 we blindfolded the child and placed him with his back to his mother about 15 feet away on the other side of the room. Since this provision eliminated all visual cues we were then careful to control the verbal ready signal given by the mother. Without warning, or preparation of any kind, the experimenter selected four words to be used for this purpose. They were "ready," "and," "go," and "all right." Sometimes I would determine which word was to be used for the whole run and sometimes either the boy or the mother would select the word to be used. The subject was never told whether he was right or wrong in his calls, as he had been up to this time, and a complete five-card, right or wrong, record was kept of his calls (see fig. 1).

In this series the mother was the only one who saw the cards before they were called by the subject. As each card was called it was turned so the experimenter could see it to get a record of the right and wrong calls.

Out of 27 runs with this method there were 373 correct calls, or an average per 25 of about 13.8.

The next series was taken on March 6, 1938, during which 14 runs, one of which was perfect, gave an average of 21 correct hits per run.

The only signal used was "ready," spoken by the mother, with a rare prompting of "go on" or "all right" when the boy failed to respond within a second or two. During this whole series a double record was kept, one person keeping a record of the calls, right or wrong, and a second person keeping a record of the cards as they were pulled, right or wrong. For the card-record there would be five checks or crosses in each column shown in Figure 1 which could be checked against the irregular call-record similar to that illustrated. For part of the recording, the two scorers sat so as to see the cards as they were drawn and before they were called, and part of the time they saw them only after the boy had given his call. For this variation in method there was no corresponding difference in his scoring which might be attributed to ESP reenforcement because of two or three senders instead of the usual one. The two scorers sat in such a way that they could see the cards as soon as they were pulled in order to facilitate the scoring. The subject could go as fast as he desired and he frequently did call very rapidly. As soon as his call was given a new card was immediately drawn by the mother and concentrated upon. During most of this series the cards were either shuffled or cut by the experimenter. The subject sat about 12 to 15 feet distant either blindfolded or in such a position that it was impossible for him to see the cards. Both scorers and the mother sat at a small card table in such a position that the scorers were nearer the boy than the mother was and within approximately two feet of her. In none of this series were any sounds emitted by the

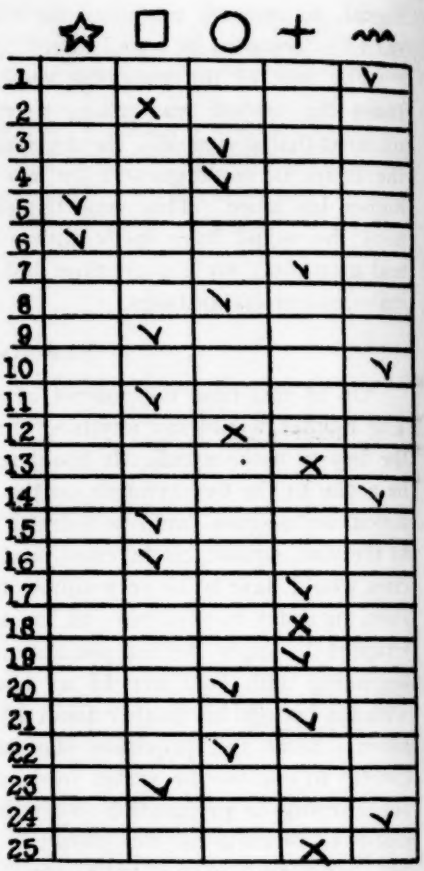


FIG. 1. Record of the fourth run on March 6. Score equals 20, with the word "ready" as signal.

sender other than "ready" or the occasional promptings already mentioned. Attempts to discover inflections which might be given "ready" either at the beginning or at the end of the word were not successful; in fact there was little change in inflection from one signal to another and I am sure there could not have been as many as five, the number necessary to account for the phenomenon.

On the fourteenth run the boy made a perfect score for which he was paid a reward of $1.00 promised by the experimenter for the first run of 25 which was perfect. Having satisfied myself that with the method employed there was little more desired in the way of high single or average scores it was decided that the time was ready to attempt some trials with the verbal signal entirely eliminated. Up to this time most of the calls had been extracted with such great effort that no attempt had been made to use anything other than the verbal stimulation of the mother. However, during most of the runs on this date there had been quite good co-operation, relatively speaking, so I was anxious to take the next step in controlling the sensory cues, especially since no one else was able to act as a successful sender.

A small hand bell was substituted for the ready signal ordinarily given by the mother. For the first trial the score was 5 correct out of the run of 25. Because the boy disliked the bell we used a pencil to tap on the table and he scored 7. For the next run we had the father say "ready" and the score was 2. Recognizing the importance of this extreme change we tried other verbal and non-verbal signals with the following results taken in order of their occurrence:

Holding up the card with the boy sitting on the other side of the room gave a score of 4.

"Go" spoken by the mother gave scores of 11 and 9 respectively.

Holding up the card again for two runs gave scores of 5 and 5 respectively.

"Ready" spoken by the mother gave a score of 13.

It was evident that for even the usual verbal signals of "ready" and "go" he had materially dropped off in his ability to score and that for non-verbal signals his scores were no better than chance expectancy. One reason may have been that he was tired from 90 minutes of card calling already and another may have been that he had been led to believe that after one perfect score we would not bother him any more that day, or he may have been confused by the change of method. It could also be suggested that, since the slight change of method, from verbal to non-verbal signals or by having no auditory contact between mother and son, reduced the scoring to no more than chance alone

would give, here was direct evidence of the existence of auditory cues of some nature. In view of the conditions and the results of the experiments on March 29, however, this appears not to be the only possible answer.

On March 29, 1938, we first secured a series of 11 runs with the verbal "ready" signal for which the average score was 11.36, about half what it had been the previous session. Wishing to eliminate entirely the auditory contact between the agent and the percipient a mechanical cricket or snapper was substituted for the "ready" signal. The cricket was operated by the experimenter because he wanted to control this signal in every way. The conditions of blindfolding, seating arrangement, double scoring were the same as on March 6. The average score for 23 runs was 7.61 or a critical ratio of 6.1. Although this average is far lower than most of those obtained previously it is still extra-chance and significant; interpreted in terms of probabilities there is not one chance in a billion that this average is due to chance alone.

The difference between the scores obtained with the verbal and the non-verbal ready signals during the same session is probably due to the lack of attention-getting value in the cricket. The mother, with the emotional tone of her voice, is able to keep the child working at the task with less distraction than is possible otherwise. The significant thing is that, with the conditions of this last experiment ruling out all sensory contact between mother and son, the score is still high enough to make some explanation other than chance necessary. Although the method of exclusion does not really prove the existence of extra-sensory perception in this case it does make this interpretation at least possible.

From March 22 to April 5 the mother kept a record of the runs made at odd times when she and the boy would run the cards. For 54 runs using the "ready" signal the average was 11.2 and for 108 runs with the cricket signal the average was 10.2. Here it will be noticed that the verbal signal is superior to the non-verbal over a large number of runs and for a fairly long period of time, confirming the results we observed on March 29. If we can take this record as having any value in itself—it may be questioned because the conditions under which it was obtained were not scientifically observed—it is extremely significant.

On April 13 I had time to observe only a few runs. The conditions were essentially the same as before with the exception of blindfolding the boy, since sometimes he objected and it was quite evident by this time that vision was not a factor related to success in his scoring, and the person keeping the call-score did not see the cards. The average

for 6 runs was 7.5 with "ready" and for 9 runs it was 5.11 with the cricket. By any method he had fallen off materially and non-verbal signals no longer gave significant results.

The last time the boy was tested was on May 20, 1938. With the use of the "ready" signal together with the call-scorer not seeing the cards, and an electric fan as a covering noise, the average for 6 runs was 8.83, or a critical ratio of 4.6. It was evident that he was not working well and the occasional trials taken at home by the mother and son alone were almost exactly the same. While the cause of this general decline in scoring can be only speculated upon, it is interesting to note, in the light of the observations by Riess,[1] concerning a hyperthyroid subject, that the subject's thyroxin dosage had been doubled a short time before.

During this same session on May 20, 15 runs were made with the cricket, this time operated by the father who did not see the cards at any time, the fan produced a covering noise, and the call-scorer did not see the cards. The average score was 4.86, or no better than chance expectancy. This lower scoring for non-verbal signals is in general agreement with previous observations and might be expected when even the more familiar method yielded only 8.83 during the same session.

We hope it will be possible to continue the investigation using only rigid control conditions. In order to make this additional work justifiable, conditions must change so the boy is able to score higher than has been the case recently.

INTERPRETATION

Several interpretations of the results are possible depending upon the bias of the judge, the weight to be given any particular experimental session, the weight to be given to chance in such card calling tests, the adequacy of the experimental controls, and the confidence in the experimenter. It is possible to accept the evidence as proof of extra-sensory perception in a particular case, it is possible to reject any or all of the evidence as inadequate, or an attitude of suspended judgment may be adopted. Taking the case as a whole I shall attempt to summarize all evidence for and against the extra-sensory hypothesis.

The first striking feature is the fact that only the mother can act as a successful sender. All attempts made by the experimenter were failures as were those made by members of his family. The father has attempted to act as agent at various times but usually the results are those to be expected from chance. It has been reported that with words,

[1] Riess, B. F. A case of high scoring in card guessing at a distance. J. PARAPSYCHOL., 1, p. 262.

objects, or ideas the child has been successful with playmates or with a few adults but it has not been possible for me to verify these reports.

The next question will likely be the possibility of minimal cues of various kinds, conscious or unconscious, which are not detectable by the ordinary person; they must always be considered in cases of this sort. All visual cues have been controlled adequately with blindfolding, placing the recipient on the other side of the room with his back towards the sender, and at all times keeping the cards out of range of his vision; so we need to concentrate upon the auditory cues. In the first place, careful listening on the part of the writer, and others, in close proximity to the sender (within two or three feet in all series and always closer than the percipient) has failed to disclose any sound other than the customary ready signal. Furthermore the child does not have extra keen audition as measured by "watch" or "whispering" tests. In addition to this the possibility of all such minimal sounds was eliminated with the covering noise of an electric fan for the verbal series on May 20 during which a low but significant average of 8.83 was secured. Extraneous sounds such as coughing, scraping of the feet, tapping with the foot, or timing in giving the ready signal have never occurred in any of the series. All suggested minimal cues have been eliminated in at least one of the significant series so they could not be the sole basis for the unusual perception. The phenomenon has been constant, though diminishing in strength, throughout all verbal series since January 30.

The third consideration is that of the possibility of a cue carried on the verbal ready signal. This could be of the nature of emotional tone, intensity, or various inflections at the beginning or end of the word or phrase. It is conceivable that five different inflections given in the verbal signal could be separate cues for each of the five ESP cards. It is, however, difficult to conceive of such obscure cues being learned by the recipient in very short periods of time. It will be remembered that on February 13 the experimenter selected the verbal signal to be used and then watched carefully for possible inflections. Also the difficulty of giving similar inflections to such different words as "and," "ready," "go," "all right," not to mention the numerous spontaneous phrases such as "what's the next card?", "go on, Bo," "let's get going," or "what's this one?" employed in the early part of the investigation, seems quite obvious. When it is recalled that the percipient cannot even do first and second grade work in school, that he cannot spell words more complex than "boy," "car," or "house," that he has an I.Q. of about 55, that he has no special memorial or other mental or

skilled ability, or that he can only with difficulty draw such simple figures as squares or crosses, it will be seen how gratuitous it is to impute him with the ability to learn numerous sets of five auditory inflections with the corresponding ESP symbol for each.

It would seem that the runs with the cricket signal make it unnecessary to question the absence of minimal sensory cues. Since the critical test in this case is the series in which an average of 7.61 was obtained, with no sensory contact between agent and percipient, the main question is how significant is this score. For this reason the actual scores with their frequencies are given.

Score:	3	4	5	6	7	8	9	10	11	12
Frequency:	1	1	1	4	5	3	3	3	1	1

The probability that chance accounts for the above scores is too remote to accept. By the approved procedure for interpreting average scores the critical ratio amounts to 6.1, or less than one chance in a billion that chance is the cause. If this series stands up under all conceivable exceptions which may be raised it would then seem reasonable to accept all the other runs which would then remove all doubt as to the extra-sensory nature of the perception.

With regard to this particular series it has been suggested that since the experimenter saw the cards at the moment they were pulled and before the subject made his call and since he was operating the cricket he might be giving some signal. The cricket was the type frequently used by elevator starters and purchased in toy departments for about 1c each and made two clicks, one when pressed and another when released. In order to advance this explanation it would be necessary to admit the possibility that the subject and I spontaneously hit upon a set of five signals corresponding to the five symbols on the ESP cards and perfecting their use in the space of time necessary to make the above 23 runs. Since we had never used this method before, such an explanation seems scarcely plausible.

It would seem that from the foregoing it is unnecessary to consider the possibility of collusion. However, I believe the agent to be wholly sincere, and I noted nothing to suggest that in any of the series there has been any deception. Even though the agent were clever enough to change methods each time the controls were changed it is not reasonable to suppose that the percipient could make this change.

CONCLUSIONS

An eight-month study of an eleven-year-old subnormal boy who allegedly possesses telepathic ability is described. The phenomenon is

only elicited with the mother as sender and occasional test trials with clairvoyance were negative. Best results were secured with any form of verbal ready signal, given by the sender, but significant results were also found with a mechanical cricket signal operated by the experimenter with no form of sensory contact between sender and percipient. A summary of results is given in Table III.

TABLE III

SUMMARY OF TEST RESULTS

Date	Average per 25	Number of runs	(calls)	Signal for next call
January 5, 1938..	6	2	(53)	Any spontaneous phrase
January 30, 1938..	18.3	3	(90)	"
February 1, 1938..	16.5	23	(595)	"
February 13, 1938..	13.8	27	....	One word "ready," "go," "and," or "all right"
March 6, 1938..	21	14	....	"Ready"
March 29, 1938..	11.36	11	....	"Ready" or "go"
	7.61	23	....	Cricket operated by experimenter
April 13, 1938..	7.5	6	....	"Ready"
	5.11	9	....	Cricket
May 20, 1938..	8.83	6	....	"Ready"
	4.86	15	....	Cricket operated by the father

Various readers will of course draw their own individual interpretations of these results. My own intimate acquaintance with the child, and the success of the March 29 series which was wholly under my control, suggests to me that explanation in terms of some extra-sensory perception of a telepathic form is favored. I realize, of course, that this is not necessarily the only explanation because the method of exclusion simply eliminates various possibilities and does not directly test the hypothesis under consideration. However, no better hypothesis seems to be available and, until it is, some hypothesis is more desirable than no explanation at all.

ESP TESTS WITH ENCLOSED CARDS

J. B. RHINE

ABSTRACT: This report covers 124,275 trials made in ESP tests in which the cards were enclosed in sealed opaque envelopes and boxes. The total score is such as would be expected only about once in 10^{11} times.

About one-third of these tests were checked immediately as usual, and these gave a critical ratio of 13.5. About half of the total trials, however, were made with delay in checking because of geographic distance, and these gave a negative critical ratio of 2.7. Delay in checking appeared inhibitory. This decline effect appeared in six of the seven test procedures used in the research.

The comparison of enclosed card tests (immediately checked) with screened gave results in favor of the enclosed. Enclosing at least offers no physical barriers to ESP, and effectively excludes sensory cues.

A tentative hypothesis is suggested to account for the scoring below chance with the delayed checking and enclosed cards.

INTRODUCTION

IN MEETING the primary requirement of excluding sensory cues in ESP tests, one of the first measures to occur to the experimenter, naturally, is that of enclosing the object to be perceived in a sealed package.

This was the precaution taken in the very first experimentation on the ESP hypothesis at Duke University. But it had long before been done in the first systematic laboratory studies of clairvoyance that were made, those of Professor Charles Richet, French physiologist, in the investigation of his subject, Leonie. In Professor Richet's tests, he enclosed playing cards in supposedly opaque manila envelopes and sealed the envelopes. His results were positive, but unfortunately we cannot on the basis of the report be entirely satisfied with the evidence for the complete opacity of the envelopes. No additional safeguards were added to reduce the possible translucency of the envelopes.

Following Richet's work, sealed envelopes were used in a number of later experiments with persons supposed to possess telepathic or clairvoyant ability. An outstanding case was the test given the Polish engineer, M. Ossowiecki, by Mr. Theodore Besterman, of the Society

for Psychical Research, London,[1] in which a remarkable success was reported.

Likewise in a series of investigations made by Mr. H. F. Saltmarsh[2] of a British medium, Mrs. Elliott, objects enclosed in sealed boxes were used as the basis for the test. In this case, too, the results were reported as successful in the representation of knowledge not available through the senses.

Sealed opaque envelopes enclosing playing cards were used in the tests for clairvoyance conducted by Besterman, Soal, and Jephson,[3] of the Society for Psychical Research, in 1930; but these tests yielded only chance results, in contrast to extra-chance positive scores obtained by Miss Jephson with open cards.[4]

There was, then, no novelty whatever in the introduction of the sealed envelope method of concealing the test object in the early research at Duke, and no reason to regard the procedure as doomed to failure. The results obtained, however, were practically those expected by chance. But during the period when this method of concealing the cards was used, the tests were administered to classes or groups of subjects and were also given (after the manner of Besterman, Soal and Jephson) in very short series of five trials per subject. Whether or not the failure to score above chance was the result of the group testing or of the short series, rather than of the fact that the cards were concealed, was not then determined. Whichever of these conditions, if any, was inhibitory to ESP, at any rate investigation along these lines yielded very little beyond chance results.

At the same time some tentative explorations made with open cards with individual subjects and in longer runs yielded results which were higher in average. While it is now easy to ascribe the difference in results in the two cases to other conditions than that of the enclosing of the cards, it was at that time regarded as possible that there was some inhibitory effect upon the subject due to his having no actual contact with the card he was attempting to call; and so for the time the enclosing of the cards was discontinued.

Somewhat later, when one of the principal subjects was located at some distance from the laboratory, it became necessary to find some method by which we could continue the tests in spite of the distance. The most suitable method seemed to be that of sending him a pack of

[1] *Proc. S. P. R.* 1933, XLI, 345-352.
[2] *Proc. S. P. R.* 1929, XXXIX, 47-184.
[3] *Proc. S. P. R.* 1931, XXXIX, 375-414.
[4] *Proc. S. P. R.* 1928, XXXVIII, 223-71.

cards enclosed and sealed in opaque envelopes, and asking him to call these and return the record of his calls for checking at the laboratory, where a code was kept. The series of tests made with this subject, with the use of these enclosed cards, has been reported.[5] The results gave the number most expected by chance; that is, the average was exactly five hits per 25, in 300 trials.

With this failure of an otherwise successfully scoring subject, the use of the enclosed card method suffered another set-back; and it was some time before it was again introduced. The next attempt, made at the laboratory approximately a year later, with another subject who likewise was scoring successfully on most of the other procedures, also resulted in failure. The score was not appreciably above mean chance expectation in a series of 102 runs through the pack. Soon thereafter another test series of 50 runs made with another outstandingly successful subject led to a considerable deviation on the negative side. Although this was interesting in itself, our experience on the whole with enclosed cards was discouraging in its failure to yield results compatible with the tests made with open cards. If our tests were a fair measure of ESP, then ESP apparently did not work with enclosed cards as we used them.

There were, however, some very good grounds for supposing that the reason for this situation was entirely mental rather than due to the material enclosing the card. A brief summary of the actual safeguarding procedures that had meanwhile come into use in the laboratory will, I think, make this point clear by showing that even more generally obstructive physical barriers had been used already with success.

First, it will be recalled that in the earliest test procedure used with the open cards (BT) the top card of the inverted pack was to be called before it was removed. If the back of the card did not give sensory cues, its inverted position represented an effective concealing of the symbol on the face of the card. The cards used were made of heavy cardboard opaque to a hundred-watt light bulb. For all practical purposes the symbol on the card when lying face down was as good as enclosed in an opaque envelope.

A further step in concealment was taken in the DT procedure, in which the cards were left in position in the pack until the entire 25 calls were made. Thus, with the exception of the top card, only the edges of the cards were visible, and the symbols were thoroughly concealed by an increasingly thick screen as they lay further down in the pack.

[5] Rhine, J. B. Extra-Sensory Perception. Boston: Bruce Humphries, 1934, p. 66.

A still greater step in concealment of the card was made with the use of opaque (usually wooden) screens, behind which the cards were placed while the subject was asked to call them.

The cards were then removed to another room from that occupied by the subject; and finally, in some experiments they were removed to distances involving hundreds of yards and eventually hundreds of miles. With walls and larger distances between the percipient and the object perceived, it is obvious that the conditions offer much more physical obstruction than anything that obtained in any of the enclosed card tests.

It seemed clear, therefore, even at the time, that the limitation in scoring when the cards were enclosed was due to a mental inhibition rather than a physical one. This mental inhibition accordingly presented a problem in itself; and this paper is a result of the investigation of it. The first step was necessarily to ascertain whether or not it was possible for subjects under any circumstances to call, by ESP, cards enclosed in opaque sealed envelopes, and, if so, to relate the results obtained in such an inquiry with the failures mentioned above, and of course if possible with normal open card test results as well.

With a slight anticipation of the substance of the report, it may be added that the first results of this renewal of the use of the enclosed cards led to considerable complication. In some instances subjects who had been scoring well above the chance level with open cards fell definitely below with enclosed, while others did not. Still others dropped, but only dropped to the chance level.

There was, however, considerable variety of conditions used in this exploratory attempt. The variation of conditions and technique did not at first permit of sufficient analysis to enable an understanding to be reached as to the probable factors leading to the complication. In fact, it was not until a considerable portion of the work reported here had been done that a plausible hypothesis for the interpretation of the unusual results was arrived at. It was discovered at length that the tendencies to negative deviation can be grouped under one condition while tendencies to positive deviation are largely confined to the opposite extreme of the same condition. The result is that at least a working hypothesis is offered to explain the earlier mystery of the difficulties with the sealed pack tests. Of its adequacy the reader must judge for himself.

PROCEDURES AND CONDITIONS

The results reported in this paper represent all of the work done with enclosed cards sent out from or used in the Parapsychology Lab-

oratory at Duke University, with the exception of that to be reported by Miss Margaret M. Price later in this JOURNAL.[6] The period covered by these tests extends from 1932 to the present year, although the greater part of the tests has been made since 1935.

The subjects in these tests have in almost all cases been subjects in other test procedures in the ESP research. They include adults and children of both sexes, both blind and seeing, with a wide professional range and a geographical distribution that covers a considerable part of the country. There have been in all 175 subjects, and the experiments have been conducted under the direction of eight investigators. Six of the investigators were psychologists, one a physician, and one an engineer.

Unfortunately for simplicity in the description of the enclosing of the cards, there were several methods used during the period of several years over which this work extended. The changes occurred as a matter of improvement in convenience or added security. However, there is no strain put by the results upon the adequacy of the methods of enclosing. As will be seen later when the enclosed decks were turned over to the subject for use in the absence of an observer, the results were such as would raise no question about sensory cues. On the other hand, when the results were such as might raise the question of alternative hypotheses, in all cases the enclosed cards were handled by the subject in the presence of an experimenter. This would of course have rendered it impossible to tamper with the enclosing without detection.

Two general methods of enclosing were used, depending upon the method used in the tests: For DT purposes whole decks were enclosed in cardboard boxes, and for other test procedures each card was enclosed in a separate envelope.

Most of the boxes were fastened by passing broad strips of gummed paper completely around the box both lengthwise and laterally, overlapping the ends and writing a signature over the end of the strip. There was one instance in which the box was sealed by filling the crevice between cover and bottom with mucilage, thus completely sealing it at every opening between the two parts.

The methods using envelopes passed through a number of stages. All the envelopes, however, were made of heavy manila paper commonly used as coin envelopes. They were just large enough to hold an ESP card with some additional wrapping material. All envelopes were sealed, and both of the sealed flaps were covered by a broad strip of

[6] Miss Price's results are not available at the time of writing and are not different in bearing from the present results.

heavy gummed paper extending over both flaps. In the case of some decks this strip of gummed paper was extended all the way around the envelope.

The first pack of sealed envelopes was made up for easy convenience by wrapping the ESP card in several layers of heavy white paper, then inserting it in the envelope and sealing as stated above. This rendered the envelope opaque to the brightest light but also made a bulky envelope. The next pack was made up by putting a blank card over the face of the ESP card, inserting and sealing both in one manila envelope and inserting this in a second (larger) one which was then sealed as above described. Several packs were made up in this fashion before a third departure led to the gluing of the blank card over the face of the ESP card. The final supply of enclosed cards added to this precaution gummed strips around the pair of cards glued together, the gummed strips running both ways around; that is, lengthwise and laterally. All the above types were opaque to light from hundred-watt bulbs both reflected and direct.

When a pack of cards was mailed to a subject working outside of the laboratory, a point was made of inspecting the pack upon its return for possible tampering. No evidence of any such tampering was discovered. It is recognized, of course, that with the exercise of the fullest available skill a clever person might possibly succeed in opening the envelopes and in sealing them up again so that the opening could not be detected. However, the scores in no instance seemed to warrant the greater labor and precaution required to afford more security on this point.

Care was taken not to leave markings on the face of the envelope which could serve as the basis of associations for the subject. Such associations could under the test conditions at the worst only mislead the subject and were reported by some subjects to be disturbing. If packs became badly worn or marked, they were exchanged.

An identifying number was placed on the back of each envelope, and a code of the symbols was for most packs taken and kept in a code book which was locked in a filing case in my office, the key to which was kept in my possession.

In certain experiments the experimenter was required to memorize the code and to make the check-up immediately following a run. When this was done, either a screen was interposed between the subject and the cards during the checking process, or else the subject was sent out of the room or (in a few cases) required to turn his back completely while the checking was done. The checking was in all respects done

in such a way as to prevent any possibility of the subject's learning the code.

There were three general conditions as to immediacy or delay in checking. There were, first, tests given in which the calls were checked immediately by the experimenter. Second, there were series in which several runs were made before the check-up. And third, there was a large amount of work in which there was considerable delay in the check-up due to the fact that the calls had to be mailed in to the laboratory for checking there.

Under the last two conditions the individual scores for each run were not given the subject. Instead, the average or the total deviation alone was given for a number of runs; and in some series not even the amount of the deviation was mentioned. The results in those cases were simply characterized as "very interesting," or "very good," or in some such language. The direction of deviation was not given. These precautions were taken to render impossible any inference about the code that might be made by any systematic study of the numbered envelopes with such design in mind.

Similar precautions were taken to prevent subjects from identifying cards in the work done for immediate checking. As mentioned above, the subject was not allowed to see the checking process. In addition, the envelopes were kept face upward during the run, which meant that the code number and other identifying marks were not visible. Also, a large portion of the subjects were blind. While many of these were not totally blind, their vision was not capable of perceiving the small symbols on the backs of the envelopes. In spite of their blindness, however, the same precautions were taken.

All the data capable of recheck (that is, of which records of both cards and calls are available) have been rechecked completely or else large enough samples have been taken to insure safety on the question of errors in recording and checking. The extent of these rechecks will be specified following the presentation of results.

Results

Totals. The report covers a total of 4,971 runs of 25, or 124,275 single trials, representing the entire range of tests to date in which ESP cards were enclosed and sealed. A total positive deviation of 947 was produced, and with a standard deviation of ± 141.0 this deviation gives a critical ratio of 6.7. It is evident that with this critical ratio, which represents roughly odds of 10^{11} to one that such a deviation might be

expected by chance, there can be no question of the statistical significance of this work taken in its entirety.

If, then, this evidence is sufficient to warrant the conclusion that something other than chance is operative, and the conditions have been such as to exclude sensory cues, we have in this work further evidence of extra-sensory perception. It is justifiable, then, to analyze further the work represented by these totals.

Scientific method requires that we attempt to associate variations in measurable results with objectively defined conditions under which they were obtained. This will lead us of necessity to deal with sub-series of this total group and to tentative groupings of results with the conditions under which they were obtained. Those apprehensive of unwarranted statistical segregation of data may be comforted, if not by the legitimacy of the analysis itself, at least by the *statistical significance of the total results.*

The following internal comparisons are of interest in these results: First, a comparison of the effect of different periods of delay between calling and checking; second, a comparison of the results obtained with the various techniques of testing, as, for example, matching compared with calling methods; and third, the comparison of scores made with enclosed and with unenclosed cards.

Immediate versus delayed check. The total number of trials falls naturally into three divisions representing different periods of delay in checking. When this division is made it is found that very marked differences in scoring level occur. All those in which the checking was done immediately following the completion of each run are regarded as Division I. Those in which the checking was delayed until the completion of a series (usually five runs; in some instances this involved, however, a delay of some hours) represent Division II. And those in which the results had to be sent in by mail and therefore the checking and reporting to the subject involved a delay of several days, fall naturally into Division III.

Each of these Divisions has a large block of data: Division I, 1,632 runs; Division II, 1,091; Division III, 2,248. *The averages decline as the delay increases.* They are, respectively, 5.69, 5.07, and 4.88. The critical ratios give the very high figure of 13.5 for Division I and the insignificant one of 1.1 for Division II. For the negative deviation of Division III the barely significant figure of 2.7 is obtained. In Table I it will be seen that the individual experimental series run uniformly positive in direction of deviation in Divisions I and II, whereas in

TABLE I

COMPARISON OF IMMEDIATE CHECKING WITH DELAYED

Subject Groups	No. of Subjects	Experimenter	Runs	Dev.	Av.	C. R.
Division I. Immediately Checked						
1. School for Blind	66	Price and Pegram	778	747	5.96	13.20
2. Duke University	17	Rhine and Associates*	251	117	5.47	3.63
3. Montgomery	2	C. H. Rice	223	169	5.75	5.54
4. Columbia	1	J. G. Pratt	222	73	5.32	2.40
5. Grand Rapids	7	E. P. Gibson	118	2	5.02	0.09
6. Oxford Orphanage	13	Pratt and Price	40	24	5.60	1.87
Total of Division I			1,632	1,132	5.69	13.50
Division II. Brief Delay						
1. Duke University	23	Rhine and Associates*	1,039	53	5.05	0.82
2. Grand Rapids	2	E. P. Gibson	52	23	5.41	1.56
Total of Division II			1,091	76	5.07	1.13
Division III. Long Delay						
1. Grand Rapids	20	Rhine and Gibson	619	—103	4.83	2.03
2. New York University	2	Rhine and Sharp	338	+ 31	5.09	0.83
3. Columbia University	1	Rhine and Pratt	335	— 5	4.99	0.13
4. Duke University	19	Rhine and Associates*	379	— 73	4.81	1.87
5. Tarkio College	9	Rhine and MacFarland	365	— 7	4.98	0.18
6. Montgomery	3	Rhine and Rice	212	—104	4.51	3.51
Total of Division III			2,248	—261	4.88	2.69
Grand Total of Divisions I, II, and III			4,971	+947	5.19	6.72

*Chiefly C. E. Stuart, J. L. Woodruff, Margaret H. Pegram, Sara Ownbey Zirkle.

Division III the trend is preponderantly negative for the individual series.

Comparison of test procedures. Most of the test procedures were used in all three of the Divisions just mentioned (I, II, and III), and therefore the comparison of results with the various test procedures can be made in all three Divisions (I, II, and III) and bear out further the effect of delay in checking. It will be seen, for example, that in the main the decline in score level with delay in checking runs through the various subdivisions representing the results of different test procedures. These results are shown in Table II. For example, the scores

TABLE II

COMPARISON OF AVERAGES FOR DIFFERENT TEST PROCEDURES

Division	Checking	No. of Subjects	Number of Runs of 25 and Average Hits for Principal Procedures* Used													
			OM		BM		STM		BT		DT		PDT		DM	
I.....	Immediate..	106	656	5.6	163	5.8	350	5.6	355	6.0	42	5.6	16	5.4	50	5.4
II.....	Brief delay .	25	159	5.2	28	4.8	...	...	123	4.9	120	5.2	12	5.5	647	5.0
III.....	Long delay .	54	888	4.8	23	4.1	344	5.0	248	5.0	467	4.9	32	5.4	246	4.9

*See glossary for brief description of procedures, except PDT and DM. PDT is pre-shuffle card-calling (See *Journal*, II, 1, 47). DM is ESP shuffle or deck matching (II, 2, 120).

for the OM procedure (see glossary) decline for Divisions I, II, and III as follows: 5.6, 5.2, 4.8; for BM, 5.8, 4.8, 4.1. The only outstanding exception to this is in the PDT (pre-shuffle card-calling, test for precognition) in which the numbers of trials are small for comparative purposes. In general, then, it may be observed that the decline with delay in checking is a well distributed effect not confined to any special test procedure.

The comparison of the results of the varying test procedures themselves is a point of primary interest. It may be seen that there is not a high order of difference shown, the greatest difference appearing again with the smaller numbers of trials. But if we omit from present consideration all groups of below fifty runs, the averages within a given Division are fairly close together; here again, too, the three Divisions differ as markedly as before.

While the BT method shows the highest average score, the next highest average is shown by a matching procedure, BM. Due to the fact that the various test procedures used were often chosen by the subject himself in order to provide conditions more satisfactory to him, the various procedures have not been equally represented. This cannot in any case represent then, a thorough and final comparison of these procedures; and no attempt is made to present further comparative statistics. These results do, however, serve the point of general indication of relative success with different procedures; and the point already mentioned, of showing that no single type of error involved (let us say, in recording or checking), would serve to account for all the deviations obtained with a variety of procedures.

Comparison of enclosed and open cards. This comparison is again an incidental one and cannot be made with nice precision. The results obtained with the enclosed cards, however, represent so large a sample that they may profitably be compared to work reported with the same

subjects and experimenters done with ordinary unenclosed cards, even though there was not an exact number of trials done by each subject for each condition. The general averages at least are of interest and value. For purposes of closer comparison we shall use only the results obtained from open card tests in which the cards were entirely screened from the subject or were separated from him by great distance. This will put the two sides of the comparison on an equal footing so far as the exclusion of sensory cues is concerned.

It is necessary to point out, too, that the work with the open and with the enclosed cards was not done in all cases during the same gen-

TABLE III

COMPARISON OF ENCLOSED WITH OPEN CARDS

	Enclosed Card Tests		Open Card Tests	
	Runs	Av.	Runs	Av.
1. School for Blind	778	5.96	911	5.57
2. Montgomery	223	5.75	158	6.04
3. Columbia	222	5.32	312	5.60
4. Grand Rapids	118	5.02	548	5.19
5. Oxford Orphanage	40	5.60	771	5.19
Totals	1,381	5.7	2,700	5.4

eral working period. There is, of course, likely to be a considerable variation on this account. The results are summarized in Table III.

It will be seen that we have no results for comparison with Division III of the enclosed card work, due to the fact that there are no tests comparable for this condition with the use of unenclosed cards. This would have required having unenclosed but screened cards called and the results sent in to the laboratory, with consequent delay.

Division II has no substantial series for comparison in open card tests. A total of 86 runs averaging 5.16, however, was made, alternating with enclosed card tests; but in these runs the cards were not adequately screened.

The comparison of enclosed and open cards, then, is confined to the immediately checked group or Division I, where there are, as mentioned, still many features of inequality in the conditions represented. However, there are 2,700 runs with cards unenclosed-but-screened available for comparison which meet the requirements. They were conducted by the same experimenters with the same groups of subjects. The average for these tests with open cards is 5.4, as compared to 5.7 for 1,381 runs

on the enclosed cards made by the same experimental groups. This difference of 0.3 per run is significant. Greater detail is shown in Table III. One group (No. 5, Duke) of the enclosed card work is omitted because no corresponding open card work was available for comparison.

DISCUSSION

Further evidence of ESP. The primary requirement of statistical significance is met by the results of this investigation. The total 124,275 single trials give a deviation that has a critical ratio of 6.72.[7] The counter-hypothesis of sensory cues is particularly well ruled out by the fact that every card was enclosed in materials unquestionably opaque to any light available to the subject.

A further question should be raised here; namely, whether the conditions excluded the possibility of the subject memorizing the code and identifying the envelopes. The question concerns two different situations: First, that in which the subjects were located geographically at great distances from the laboratory where the code was kept locked up. It would have required collusion with the laboratory staff to obtain the code. One is assured further by the fact that the score averages from all such subjects have been either below the mean chance expectation or very close to it. (See Division III, Table I.)

The other aspect of this question concerns subjects who were tested in the presence of the experimenter. In this instance the only opportunity for the subject to acquire a knowledge identifying the cards in the envelopes lay in his finding visual cues on the front of the envelope and then observing some give-away indication of the experimenter during the checking-up process by which he would identify that particular envelope. Even granting the presence of some identifying visual cues on the envelopes, the other requisite for the identification of the card was definitely lacking, since in the checking-up process the subject was not allowed to look on. In the case of the blind children it can be seen that this alternative too is ruled out by their deficiency as well as by the experimental conditions.

It is next in order to consider the adequacy of the safeguards against errors in handling the data. We find that particularly with the more recent work the evidence is quite reassuring against this hypothesis. First, all of the delayed check work of Division III, except fifty runs, was independently recorded and was capable of independent recheck-

[7] The only enclosed card work done to date not included in this total, that of Miss Price already mentioned, will increase this critical ratio to a still higher figure.

ing. The critical ratio of 2.7 for this block of data is based, then, upon this precaution.

A sample of 433 runs was checked from the original codes. (The average from Division I and II was raised from 5.23 to 5.24; errors of omission are the more common.)

It was possible to recheck 320 runs of the Price and Pegram work with the blind by confirming the arithmetic and making certain that there appeared in the card column for each run exactly five of each symbol. These scores averaged 5.88. Since this average is very close to the total average of 5.96, the sample may be considered an adequate check upon the total results.

Further reassurances on the safeguarding of the checking procedure are as follows: The short series with the Oxford Orphanage group shown in Table I which averages close to the average for Division I was conducted by two experimenters so that double witnessing prevailed throughout. This was true likewise of a unspecified part of the work at the School for the Blind. A further reassurance is offered by the fact that, as Table II shows, the averages for the matching and the calling methods are approximately similar. In view of the very different methods of checking and recording, this would be very unlikely, on a basis of supposition of error, to account for the deviations. Finally, it should be said that in those matching methods for which no independent records of calls and cards were kept, we have eight experimenters responsible for the recording and checking. In the matching procedures the checking is a matter of very simple counting of the number of hits. While error is possible, it is much less technically probable than that of the task of the average physical scientist engaged in reading various balances and meters. Probably in all such instances error occasionally occurs. It cannot, however, reasonably be supposed to have become a habit for so considerable a number of trained professional people as are responsible for these tests. Nevertheless, the reader will have to select according to his own criteria what standard he will accept; it is likely that in some portion of the data he will find that it has been met.

Delay in checking. It is definitely indicated by the averages of Divisions I, II, and III in Table I that a falling off in score average occurred with delay in checking. That this decline is not a mere accident is indicated in a number of ways: First, by the critical ratios of the different subdivisions (these latter were made entirely on the principle of delay in checking); second, there is the fact that, as shown in Table II, even when the test results are further subdivided according

to the test procedure used, there is still shown in all but one subdivision the same decline with delay. It is noteworthy that in the main the subjects who had been scoring high in open card tests and who scored well in the immediate checking dropped lowest in score level in the delayed checking. This can be seen by a comparison of the Montgomery group for Divisions I and III. Also, the lowest scoring subject, H. P., in the Duke University group for Division III was one of the outstanding subjects in tests with open cards. The Grand Rapids group in Division III furnished two outstandingly low scorers. One of these, L. H. G., was the principal high contributor in Gibson's work reported in the December 1937 JOURNAL.

It seems clear that there is some peculiar mental inhibition operating in suppressing the scoring in this delayed check, enclosed card work; that is, if the evidence is sufficient to warrant some speculation. First, it should be noted that mere delay in itself can hardly be regarded as the whole explanation, in view of the fact that, as I reported in this JOURNAL in September of last year, in 101,450 trials conducted with the subjects a long distance from the cards, a significantly positive deviation was obtained in spite of a delay equal to that of the present work with enclosed cards.

But what difference may be supposed to inhere between the distance work with the cards far away from the subject, and the present work in which the cards are enclosed in sealed envelopes and are in the possession of the subject? In both cases the call series has to be copied down and sent to an experimenter by mail and a reply awaited. Both methods presumably are less interesting, or at least less immediately satisfying than the method of immediate checking.

If there is anything about the enclosed card method of delayed checking that will account for the results obtained, it would have to be something which not merely lowered the scoring to chance, but which drove it into a negative deviation; that is, introduced a reversal of the tendency to hit the right card and caused a tendency to be brought into play to miss the cards more often than chance would allow. Presumably this contrary urge or negativism was in all cases entirely unconscious, though on this score the exigencies of courtesy might well have kept a dislike for the enclosed card, delayed check methods from being expressed. One thing is clear, however, that it was not easy to get subjects to take part in any appreciable number of runs with the enclosed card, delayed check method.

But if we suppose a negativistic attitude toward this procedure and condition, what explanation can be furnished for negativism in this case

that would not apply to simple distance tests of clairvoyance with open cards in the hands of the experimenter. At present we have little more than sheer speculation to furnish a working hypothesis. But it appears probable that there is more annoyance given a subject by a sealed package in his possession of which he tries to divine the contents than would be the case if the object were simply far away from him. In the case of the enclosed cards the same envelope is given him time after time, and each time he can only "guess" what it is and is never told; whereas in the open card distance experiment the subject never has to deal with the same card-order over again. It is, I am inclined to suppose, probably the continual frustration of being repeatedly asked to call a pack of cards with no satisfaction ever given as to just how successful a given trial or even a given run has been. It seems thus probable that a positive urge can with sufficient striving, as in the fable of the fox and the grapes, turn sour and become negative. The fact, however, that the subject does not wish to give up the testing which he has promised to carry out may keep him from giving conscious recognition to his annoyance. So much, then, for a merely plausible and wholly untested view of the results of the tests in Division III.

In this connection it should of course be recalled that subjects in ESP tests have been successful in obtaining low scores by deliberate effort at missing the calls. A number of reports of this character have appeared in the literature on ESP, in general the result being that the negative deviation in the low-aim calls approximates that of the positive deviation of the same subject in high-aim calls. The average negative deviation in Division III is not as great as the average positive deviation in Division I, it is true; but even if the speculative hypothesis just proposed to account for the negative deviation in Division III be sound, it need not be supposed that all of the subjects in the tests reported in Division III were negatively oriented. It should be recalled, too, that other negative deviations have been reported without deliberate effort on the part of the subject operating to bring them about. These deviations are only suggestive, but there is a certain obvious relationship to the present work which merits mention.

The first case of negative deviation is that reported by Estabrooks in his series No. 4.[8] This series, in contrast to the other three, which were positive, gave a significantly negative deviation on the calls of the subjects for the suits of playing cards. Thirty-two runs of 20 calls each averaged only 4.06, whereas mean chance expectation is 5. Calls

[8] G. H. Estabrooks, "A Contribution to Experimental Telepathy," *B. S. P. R. Bull.* V, 1937 (Feb.), p. 19.

as to color were also negative. The difference between Series 4 in Estabrooks' report and the other three series, which were positive, is that the subjects were asked, contrary to their expectation, to wait over for this series after they had already finished one. The subjects were students at Harvard College and are represented by Estabrooks as having had no great intrinsic interest in the tests. For example, he was unable to get them back for a retest. One can readily suppose that the process of waiting around for an unscheduled repetition of tests which were not in themselves exciting would stir some at least unconscious impatience and that this might evoke a certain amount of negativism in some subjects.

A similar instance may be found in my report of the early experiments at Duke in the case of the subject, Linzmayer.[9] On this occasion I took advantage of the subject's desire to leave by pre-arranged plan, and persuaded him to stay over for a long series of tests. He did this under sheer pressure of courtesy, as was obvious. His score, which had been averaging approximately 9 hits per 25, dropped to an average of 4 for the next 16 runs.

It is a far jump, but perhaps not without some connection, to pass next to the negative deviations obtained from psychotic patients by Shulman[10] and Van Wiemokly. They found averages of 4.72 and 4.77 for involutional melancholia patients, in which classification there seems to be general agreement that the prevailing attitude of the patient would probably be one of impatience with the tests.

At any rate, it will be of interest to watch for possible cases of negative deviations under conditions in which the subject is to some degree frustrated or led to take a negativistic attitude toward the test. Perhaps some more specific experimental method can be found to evoke this attitude consistently.

Comparison of procedures. The conspicuous thing about this comparison, as has been mentioned already, is the extent of the decline with delay in checking shown by the various subdivisions according to test procedure. It is also noteworthy that the general level of scoring varies very little from one procedure to another within a given Division. In view of the fact that the subjects were as a rule allowed to select their favorite test procedure and yet that there is a substantially large sample of the principal method, it would appear relatively unlikely that any important difference lies between the methods, except as the subject's

[9] Extra-Sensory Perception. Boston: Bruce Humphries, 1934, p. 59.

[10] A Study of Card-Guessing in Psychotic Subjects, J. PARAPSYCHOL., 1938, II, 95-106.

attitude itself becomes determinative. It has been shown by a number of investigators that for given individuals certain procedures appear to be advantageous. A general perspective of these results, however, seems to suggest that these differences are entirely personal and are perhaps matters of taste.

The fact that the total number of runs differs from one subdivision to another might suggest to the critic that a series under a particular heading was discontinued when a favorable critical ratio was obtained. This, however, was not the case, as can readily be understood from the fact that these totals represent summaries of a number of independent investigations in which no investigator knew what the other had done or when the others had discontinued. The total effect, then, may be regarded as quite independent of calculation. In fact, the critical ratios not only for the subdivisions but for the total of the Divisions and of the research itself as a whole were not struck until the occasion of writing this report.

Comparison of enclosed with open cards. Little can be said as a result of this comparison, because of the lack of precise comparative conditions. The primary problem so far as the enclosed card work is concerned was to find out if significant ESP data would be given. Such comparisons as are available are largely incidental; nevertheless they do serve to show, as Table III bears out, that there is certainly no disadvantage in having the cards enclosed in sealed envelopes, provided the subject can be acquainted with his success reasonably promptly and his spontaneous interest in the task thus maintained.

In view of this comparison it is clear that there is no physical obstruction to ESP in the enclosing of the test cards; it is thus all the more certain that if there is, as appears, some reversing tendency induced with the delayed checking of these sealed packs, it must be due to a psychological barrier and not to a physical one.

Summary

A total of 4,971 runs was given a group of 175 subjects under the observation of nine experimenters; and a total positive deviation of 947, with a critical ratio of 6.72, was obtained. The conditions for the exclusion of sensory cues and precautions taken to avoid disclosing the identity of the enclosed cards, along with independent rechecks on the security of the recording and checking, leave no recognized alternative to the hypothesis of extra-sensory perception.

It was found that there was a marked correlation between delay in the checking and decline in the scoring. When the cards were checked

immediately, positive deviations were obtained; and when the cards were sent out to subjects, with the necessity of using the mails for the conduct of the tests, the total score averages became significantly negative. This does not, however, establish that mere delay in checking in itself is sufficient explanation. A tentative hypothesis is suggested to account for the depressed scoring tests involving delayed check.

Comparison of test procedures reveals no outstanding difference in favor of any of the procedures. In six out of seven test procedures used the decline with delay in checking occurred.

An incidental comparison of open card tests with the enclosed card work of the present report gives an average somewhat in favor of the enclosed card work. This incidental comparison is regarded merely as demonstrating that no insuperable physical handicap inheres in the enclosing of the card.

The suggestion is left that where there appears to be a depression of scoring as a result, apparently, of the delayed checking of enclosed cards, this is entirely due to mental rather than physical inhibitions.

AN EXPERIMENT TO TEST THE ROLE OF CHANCE IN *ESP* RESEARCH*

CLARENCE LEUBA

ABSTRACT: When a subject attains a deviation considered significantly above chance the question arises whether the score really shows some extra-chance factor operative, or whether it is an unusual but predictable chance deviation. In an experiment to study the frequency of "significant" deviations among chance data, the experimenter shuffled two decks of ESP cards and matched them 2,682 times. The scores were recorded on cards, each card being denoted a "subject." The number of runs for each "subject" was determined by the experimenter's judgment as the experiment progressed. The total average was 5.03 hits per run. Of 87 arbitrary "subjects" the total scores of three exceeded a 2.5 critical ratio, and seven were found to have somewhere within the series eight or more consecutive runs exceeding that critical ratio.

Dr. Leuba is Professor of Psychology at Antioch College.

In a discussion of "Mathematical Techniques Used in ESP Research," Greenwood and Stuart write as follows: "An anti-chance probability in favor of a particular hypothesis must be considered in respect to *all* the data of the experiment. For example, if 200 subjects were given a test of 100 trials each, and one of the scores yielded a significant critical ratio of 2.5, the conclusion that the given subject had demonstrated an extra-chance factor would be without basis. One such critical ratio at least would be expected by chance in the population of 200."[1]

That this statement is true may be demonstrated experimentally; it is only necessary to match two packs of ESP cards with each other, letting an arbitrary number of runs represent the work of a first imaginary "subject," a second arbitrary number of runs the work of a second imaginary "subject," and so on. In this way, on the basis of chance alone, it is actually possible to obtain scores yielding a significant critical ratio of 2.5 or more for many hundreds of trials for certain lucky "subjects."

A more detailed description of the experimental procedures and results follows. Two packs of ESP cards were each shuffled three times

* Presented September 7, 1938, at the annual meeting of the American Psychological Association in Columbus, Ohio. [1] J. Parapsychol., 1937, I, 224.

and cut. (This was always done immediately before each run.) The packs were then placed side by side face up. The top card was drawn with the right hand from the right pack and with the left hand from the left pack; then the second card likewise from each pack, and so on. A mental note was made each time the card in one pack matched the corresponding card in the other pack and at the end of each run the total number of hits was recorded on a 4″ by 6″ card. Approximately every five or six runs, the packs were matched a second time, before shuffling, as a check on the reliability of the scoring. (In the total of 2,682 runs made in this experiment, there was found one error of one hit.) The first 10 runs were recorded on the first 4″ by 6″ card and were supposed to represent the work of a first "subject," the second 10 runs were recorded on card number 2 for a second "subject," and so on. One exception was made to this procedure in that if a "subject" happened to score significantly before 10 runs (according to the Table for Grading Total Score in Stuart and Pratt "A Handbook for Testing Extra-Sensory Perception," 1937) we proceeded immediately to the next "subject," even though the first one had completed less than 10 runs. Having secured in this fashion a total of 87 "subjects,"[2] the scores for each "subject" being on a separate 4″ by 6″ card, those "subjects" who had averaged more than 5 hits per run, or who had gotten high scores during the last few runs, were given a second session. If, in this session, they averaged less than the chance expectation of five hits per run, they were dropped as "subjects"; in the few cases where the good scores in the first session were followed by good scores in the second session, a third session was given; if this was also good, a fourth session, and so on. After the first session, the number of runs given to any one "subject" at any one session depended upon the amount of time the experimenter had available and upon a tendency on his part to stop, in the case of a "subject," who was coming to show only close to chance scores, after a definitely poor run; and, in the case of a "subject" who was showing particularly good results, to stop after an especially good run or series of good runs. The decision as to whether to continue with the same "subject" for another run, or whether to turn to another "subject" was always made *before* the cards were given their usual three shuffles and a cut.

The 87 "subjects" made a total of 2,682 runs, and scored 13,496 hits, thus averaging 5.03 hits per run; this differs from the theoretically expected average of 5.00 by less than one S.D. Seven of the 87 "sub-

[2] In this preliminary work of "discovering good subjects," I had the assistance of three students: Carl Hamilton, James McDowell, and Catharine Robinson.

jects," however, did have, at one time or another, eight or more consecutive runs with a positive deviation above chance expectations of 2.5 or more standard deviations. For instance, "subject" F, after scoring poorly on a few runs, went through 40 runs (1,000 trials or matchings) scoring 53 hits above chance and averaging 6.3 hits per run. The critical ratio for these 1,000 trials is 4.1. In this case, in addition to selecting only one "subject" out of a larger number of "subjects," we are also selecting only a certain batch of his data, namely 1,000 consecutive trials. Both before and after these 40 runs "subject" F fell to the chance average of 5 hits per run or below.

A selection of "subjects," without any selection of batches of data within the performance of any one "subject," suffices, however, to obtain good scores. Three of our "subjects" (A, B, and E) scored significantly above chance *from the beginning.* Table I gives their scores for *all* their runs for as long as they were able to score successfully.

TABLE I

HITS MADE BY "SUBJECTS" A, B, AND E ON EACH RUN

Run	A	B	E	Run	A	B	E	Run	A	B	E	Run	A	B	E
1	8	7	6	16	7	6	5	31	*	5	5	46	*	*	7
2	5	4	6	17	6	4	10	32	*	6	8	47	*	*	7
3	9	1	11	18	6	7	2	33	*	10	6	48	*	*	7
4	8	6	5	19	7	6	4	34	*	5	5	49	*	*	3
5	11	8	4	20	6	7	5	35	*	5	3	50	*	*	4
6	5	7	7	21	9	6	3	36	*	6	3	51	*	*	7
7	9	7	7	22	7	5	7	37	*	7	5	52	*	*	12
8	7	11	8	23	5	5	3	38	*	8	5	53	*	*	5
9	6	6	9	24	1	5	5	39	*	6	9	54	*	*	7
10	6	5	4	25	7	10	4	40	*	7	4				
11	3	6	7	26	6	9	6	41	*	6	5				
12	4	6	3	27	7	5	5	42	*	*	7				
13	5	1	4	*28*	4	10	6	43	*	*	4				
14	5	9	5	29	9	6	5	44	*	*	7				
15	5	3	6	30	*	6	7	45	*	*	8				
								Average per run.....					6.3	6.2	5.8
								Critical Ratio.....					3.5	3.8	2.8

It will be noticed that "subject" A averaged 7.4 hits per run during his first 10 runs (250 trials) and that he never once fell below chance during these 10 runs (C.R. 3.7). This "subject's" record, both for these 10 runs and for the total of 29 runs is such that, according to the Handbook for Testing Extra-Sensory Perception by Stuart and Pratt (p. 41-42), the scores are "very good" and "clearly significant."

On page 59 these authors write that "If the critical ratio is between 2.5 and 3.0, you can be relatively certain that the results did not happen by chance. If it is between 3.0 and 4.0, you can be very sure."[3]

A few of the "subjects" who had scored definitely *below* chance during the first session of 10 runs were continued to see how many runs it would take before they reached the chance expectation of 5 hits per run. One of these poor "subjects," Z, during a total of 131 runs (3,275 trials) *never* reached this chance expectation; he always remained below chance. His total number of hits was 615; this is 40 below the number expected by chance.

The difference between good "subject" B and poor "subject" X is a significant one. The former (Table I) averaged 6.2 hits for 41 runs (1,025 trials); the latter 4.3 hits for 42 runs. The difference between these two averages is 1.9 hits, with a standard deviation of .45 and a critical ratio of 4.2.

DISCUSSION

Let us return to the first sentence in the statement by Greenwood and Stuart quoted at the beginning of this article: "An anti-chance probability in favor of a particular hypothesis must be considered in respect to *all* the data of the experiment." It seems to me that *all* the implications of this statement have not been fully understood by many of the experimenters in the ESP field. It implies that, using the accepted mathematical methods for calculating anti-chance values, it is *not* permissible:

1. *To select certain batches of data* made by one subject, as, for instance, the data secured between certain dates, and to use those data as evidence for ESP.

[3] Our results might be interpreted, by those inclined toward a belief in precognition, as indicating that I knew *before* shuffling each pack three times and cutting it, what the general order of the cards in each pack would be *after* the shuffles and cuts and therefore that I tended to stop using a good "subject" when I had a precognition that the packs, after shuffling and cuts would match poorly; and conversely, that I would continue with a good "subject" when I had a precognition that the packs after shuffling would match rather well. To this interpretation one might reply that I was conscious of the following factors as determining when to stop with a given "subject" at any particular session: (1) some appointment or outside interruption; (2) whenever I had gotten 8 hits or better for the last run, or, having worked for some time, being satisfied with a 7 or even a 6; (3) after two or three definitely poor runs. Furthermore, it should be remembered that I continued for a second, third, fourth, etc. session with "subjects" who had been good on the previous session or sessions. Thus, whether the good scores came to one of the subjects or to another made no difference. Quite apart from these considerations, however, it would seem safer to interpret the results from the point of view of the accepted mathematical principles stated by Greenwood and Stuart at the opening of this article.

2. *To select certain subjects* out of a much larger number of subjects tested, even though *all* the scores of these selected good subjects are used in the evidence for ESP.

3. *To select certain total investigations* out of all the investigations now going on in this country (and elsewhere) as evidence for ESP. If there are enough investigations it is conceivable that a few experimenters will get an average of between 5 and 6 hits per run, for the total of all the runs of all their subjects; because of the large number of runs involved, this average may yield a critical ratio of more than 2.5.

In other words, the large (but unknown) number of ESP runs being made at present, both inside and outside university laboratories, makes it impossible to tell what the chances are for a *purely chance* occurrence of any particular number of hits.[4]

In addition to the three types of selection mentioned above there is a second non-ESP factor which probably influences the results to an undetermined extent; namely, the tendency to continue with a good subject until he has achieved satisfactory results; or, to put it negatively, to refuse to stop until he has achieved results at least somewhat comparable to his previous performances. It is well known that a gambler with unlimited resources could refuse to stop until he was ahead by a substantial amount; likewise, if one has unlimited time and patience, one may refuse to stop in any ESP experiment until one is significantly above chance expectations. If the experimenter's and the subject's desire for success is strong enough—and it has been repeatedly stated that this attitude is essential for success—this factor might well influence the results favorably at times.

[4] The results of our experiment gain in significance when one realizes that most of the ESP research now being carried on under more carefully controlled conditions, no longer yields the earlier and astounding averages of 8 to 12, or even more, hits per run; but averages instead usually only between 6 and 7 hits, or less. It is true that these recent investigations sometimes average such values for many thousands of trials, whereas we got them for a maximum of 1,350 trials. Our good "subjects" were selected, however, out of 87 "subjects," whereas the good subjects reported in the literature represent a selection out of the hundreds of subjects being tested and the hundreds of thousands of trials being carried on at present both in and out of university laboratories. The larger the number of runs made, the greater the chance of securing, at one time or another, a series with an unusually large number of hits.

AN EMPIRICAL INVESTIGATION OF SOME SAMPLING PROBLEMS

J. A. GREENWOOD

ABSTRACT: Certain problems of sampling arising out of the ESP research are examined here in an empirical way based on the 500,000 matchings of the chance check series reported by Greenwood in the last Journal. There is found the approximate amount of revision of the critical ratio required to permit optional stopping of an experimental series, and the extreme critical ratio obtainable from optional stopping in the whole series of 20,000 runs. Also there are submitted empirical checks on the expectation of blocks of data of various lengths and given scores or score totals; and finally the frequency of a given block (ten runs) of fixed critical ratio with optional starting point in a lengthy series. Recommendation is made on the needed revisions as well as the methods of evaluation of a given block of data against a general background.

Dr. Greenwood is Assistant Professor of Mathematics at Duke University. This is the second of a series of empirical analyses made with the large check series referred to above.

IN THE first paper by C. E. Stuart and the writer (3) indication was made of the existence of ESP sampling problems over and above those treated in that article. The following report is concerned with the study of some of these problems, but it must again be recognized that there are still other sampling problems which will not be included in the present treatment. The work here submitted represents an entirely empirical study, leading to only tentative recommendations. But it is hoped that this will be the forerunner of a series of researches leading ultimately through the joint efforts of several persons to development of a code of proper procedure for the entire array of sampling problems applicable to ESP research, though by no means confined to it.

The main problem of this study is that of finding the effect of optional stopping of a test series upon the probability value of the critical ratio. To what extent are the chances of exceeding a given critical ratio altered by freedom to choose the place to stop the series? This is important to know in order to free the experimenter from the necessity to lay down in every research a fixed goal at the start. Conditions may render estimates in advance difficult. If a proper correction can be found or the extreme amount of required revision determined, the investigator has a more adaptable instrument at hand for his use.

Again there has risen frequently the question whether under probability theory there is not full certainty of obtaining a critical ratio of any given size provided the series is sufficiently prolonged. Might not some of the lengthy ESP series running into the hundred thousands of single trials allow enough play for some high critical ratios on the basis of chance alone? This question arises sufficiently to warrant an empirical inquiry as to the extremes of critical ratio obtainable in the largest chance series yet on record, 500,000 trials, by starting at the beginning and choosing the stopping point giving the greatest ratio. The length of this series of a half million is well above that of the largest ESP series yet reported.

A third problem is considered empirically here largely to make its solution more realistic and objectively apparent to a great many people at present confused about it. If the probability is expressed that a given sub-series of, say, ten runs with a given average would be expected once in 150 times, would we actually find this estimate correct on a chance series? Or might we not, it may be asked, find that with the ESP card material there are many more "significant" critical ratios than the law (of chance) allows? Again, is it easier to get a significant critical ratio with short sub-series than with long? And, finally, do the frequencies for larger score groups turn out to be good binomial fits as I found for the single run scores?

As a fourth and most tentative question, what is the frequency of occurrence of a block of data of fixed size having a given critical ratio, starting anywhere in the series? Again, the exploring experimenter wants to know how to evaluate spots or patches of apparent interest in ascertaining whether he has a chance or extra-chance occurrence. Sensitive methods for these evaluations are needed. An empirical approach to these questions has the advantage of concreteness.

Perhaps none of these problems are crucial to any particular research, but in addition to the advantage of greater precision and in the interests of securing better adjusted tools of measurement for the experimenter, it seems worth while to present these findings.

The several analyses reported in this paper are based upon the empirical series reported in the preceding number of this JOURNAL.

POSSIBLE ADVANTAGE OF OPTION IN STOPPING; EFFECT ON INDIVIDUAL SUBJECT'S CRITICAL RATIOS

The problem of this section as empirically investigated can be stated thus: What is the probability that on some run in a series between 10 and 50 runs, inclusive, a "subject" will achieve by chance a total critical

ratio of +2.5 or better? To answer this objectively the 20,000 runs of the series used were divided into 400 consecutive blocks of 50 runs each. The critical ratio of the first 10 runs of each block was found, using the standard deviation of the binomial. Then to each of these groups was added the score of one run at a time, each time noting whether or not the total deviation of that block up to that point had reached $X = +2.5$. This procedure was continued to the end of each 50-run block in order to see how many times in each block this ratio was attained, if at all. Of the 400 "starters" 11 attained +2.5 at some time. Thus the empirical probability to be associated with $X = +2.5$ in this particular case is, roughly, .028 ± .008 (S.D.) instead of .009 of the corrected normal or .011 of the binomial obtained by ignoring the effect of the sampling freedom which was introduced.

It was of interest to note the upper range of values of X in excess of +2.5. The largest critical ratio reached in any block was +3.0.

Table I gives the summary of all blocks attaining $X = +2.5$.

TABLE I

Runs	10	11-20	21-30	31-40	41-50
Attained $X = +2.5$ for first time in indicated run.....	3	4	1	2	1

In order to throw more light on the situation through larger numbers, the same thing was done for a critical ratio of +1. However, each block of 50 runs was increased to one of 100 runs by advancing 250 runs in the series and then counting off 50 runs *at right angles* to the original order of recording. For the purpose at hand this would give the practical effect of a million-card-call series. Also, as soon as a block attained $X = +1$ this entire block was discarded. The results are given in Table II.

TABLE II

Run	10	11	12	13	14	15	16	17	18	19	20
Attained $X = +1$ for first time in indicated run...	56	12	10	6	9	10	6	3	3	2	3
Run		10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Attained $X = +1$ for first time in indicated run.....		56	64	16	13	12	5	10	6	1	6

Of the 400 blocks, 161 had attained $X = +1$ (or greater) by the end of 50 runs; and of the remainder, 28 more had attained it by the end of 100 runs. If the same rate of increase of successes were assumed for Table I, extended to 100 runs, as appeared in Table II and proportional allowance made for the greater loss of successful blocks by run 50 in Table II, then the empirical probability of achieving $X = +2.5$ within 100 runs would be. 035 ± .009.

The entries of Table II show that the chances of attaining $X = +1$ for the first time are greatest on the early runs and fall off rather steadily thereafter. The empirical entry under run 10 corresponds to the probability obtained from the tables of $X \geq +1$ or $+2.5$ as the case may be. With freedom to stop the series at will, all the succeeding numbers up to that run must be added to the first entry, increasing the tabular probability proportionately. It is thus evident that where a probability as small as .0067 (approximately 1 in 150) is desired as the level of significance, under these conditions of sampling a somewhat larger critical ratio than $X = +2.5$ is required. A definite but tentative recommendation will be found in the conclusion of this paper.

EXTREME CRITICAL RATIOS FROM OPTIONAL STOPPING IN ONE BLOCK OF 20,000 RUNS

Directly connected with the above problem is that of seeing how extreme a critical ratio a single "subject" can obtain in the longest chance series available; namely, that of 20,000 runs already referred to. Again starting from the first 10 runs and following a sliding scale of deviations necessary for $X = \pm 2.5$ throughout the entire 20,000 runs, the following was observed.

On run 13 the total deviation gave $X = -2.5$, and remained at that or below (algebraically) until run 33, inclusive. At run 29 $X = -3.25$, and at run 33 $X = -2.61$. The next appearance of anything beyond the criterion was on run 12,417 with a negative deviation of 562 giving $X = -2.52$. The series continued giving –2.5 or less at intervals until run 12,499. Again at run 12,657 and for the four succeeding runs X was –2.5 or less. The next time X reached this point was on run 13,227, and it stayed below until run 13,241. The critical ratio never again fell as low (algebraically) as –2.5. Nothing worthy of comment showed up on the positive side. Excluding the critical ratios of the first 33 runs, the numerically greatest critical ratio achieved thereafter was 2.57 on the negative side.

The writer attaches the most significance to the fact that after the series reached stability (i.e., so large that the addition of a single run

would affect it but slightly), no unusually large (numerically) critical ratios were observed. However, the appearance on run 29 of an $X = -3.25$ must not be entirely discounted. Considering the likelihood of positive or negative values and the increases in probability discussed in the section above as resulting from optional stopping, it would seem to be a case of an actual occurrence whose improbability is of the order of two or three hundred to one. We may recall in this connection again the fact that in 400 blocks of 50 runs where positive deviations alone were considered, only one block attained at some time a critical ratio of $+3.0$.

FREQUENCY DISTRIBUTIONS IN A LARGE CHECK SERIES BY BLOCKS OF 5, 10, 25, AND 50 RUNS

The probability of attaining a given critical ratio for various sized blocks of runs has been answered theoretically for the pure matching case by Huntington (5) and for the binomial by Stuart and Greenwood (6). The following Table III admits the same information empirically for the above-mentioned groupings for the free-calling case of the empirical check series. The data were grouped into consecutive blocks of runs of the above sizes in turn and deviation from expected score of each block found.

Of interest is the distribution of the groupings by 5 runs compared with the theoretical distribution of the binomial $4{,}000\ (.8 + .2)^{125}$. (The latter terms were computed using tables of logs of factorials.) A chi-square test of comparison gives $p = .4$, an excellent fit. By noting the points $X = +2.5$, for example, the distributions will be seen to bear out the theoretical inferences of the Table IV of an article previously referred to (6); namely, the smaller the number of runs taken, the more likely is one to attain a given deviation ratio. In other words, the relative frequencies beyond these points increase with a decrease in the size of the groupings.

SELECTION OF A BLOCK OF FIXED LENGTH WITHIN THE DATA OF A CHANCE SERIES

Suppose the best consecutive 10 runs, say, of a given set of data were selected. Upon what basis could it be evaluated? To approach this question, all consecutive 10 run blocks from runs 4,001 to 8,000 inclusive (3,991 of them) were investigated as to critical ratio, X. The following table shows the frequencies obtained for the ranges indicated.

Of the five whose numerical critical ratio was as large as 3, none exceeded it enough to occasion mention. Taking the probability of X

TABLE III

Deviation	Grouped By				Frequencies of Binominal $4,000\,(.8+.2)^{125}$
	50's	25's	10's	5's	
33	1	..	..	..	.....
32	0	..	..	..	.....
31	0	..	..	..	.....
30	0	..	..	..	.....
29	2	1	..	..	.....
28	0	0	..	..	.....
27	1	1	..	..	.....
26	3	0	..	..	.....
+25	3	1	..	..	.....
24	3	3	..	..	.....
23	1	0	..	..	.....
22	3	3	..	..	.....
21	2	2	..	..	.....
20	4	2	..	..	.....
19	4	9	1	..	.....
18	6	6	2	..	.....
17	4	3	3	1	8.05
16	4	7	6	0	.....
15	5	6	3	2	.....
14	5	9	15	2	.....
13	8	13	20	5	6.96
12	6	23	17	10	12.02
11	12	20	20	23	19.99
10	13	15	32	34	31.98
9	6	21	44	37	49.19
8	7	23	60	64	72.72
7	5	27	67	106	103.21
6	8	28	77	119	140.55
5	13	15	81	186	183.46
4	13	28	105	219	229.32
3	11	27	94	268	274.24
2	11	26	104	337	313.41
+1	12	34	120	313	341.90
0	12	33	136	377	355.58
−1	13	47	126	340	352.06
2	11	34	136	335	331.35
3	11	36	128	323	295.96
4	19	23	114	282	250.43
5	11	36	84	206	200.35
6	13	29	80	135	151.20
7	8	19	88	104	107.40
8	10	20	61	72	71.60
9	8	19	44	48	44.67
10	10	24	40	24	25.99
11	8	20	32	13	14.05
12	14	11	22	11	7.02
13	11	11	13	2	.....
14	5	19	11	1	.....
15	4	12	4	0	.....
16	11	11	5	1	5.36
17	2	7	3	..	.....
18	4	10	1	..	.....
19	6	6	0	..	.....
20	7	6	1	..	.....
21	4	4	..	..	.....
22	8	3	..	..	.....
23	2	0	..	..	.....
24	1	3	..	..	.....
−25	0	0	..	..	.....
26	1	0	..	..	.....
27	0	1	..	..	.....
28	0	1	..	..	.....
29	2	0	..	..	.....
30	2	1	..	..	.....
31	1	1	..	..	.....
32	3	..	..	..	.....
33	0	..	..	..	.....
34	1	..	..	..	.....
..	..	..	..	..	.....
48	1	..	..	..	.....
Total	400	800	2,000	4,000	4,000.02

TABLE IV

$+2.5 \leq X < +3$	$-2.5 \geq X > -3$	$X \geq +3$	$X \leq -3$
Freq. 18	29	2	3

outside the range of ± 2.5 as .018[2] one computes that on the average 2,890 *independent* groups of 10 runs would produce 52 critical ratios numerically as large as above. This is comparable to the above 3,991 *correlated* ones which produced the same number, 52. It may be that later theoretical investigation of this problem will produce a practical method of evaluating a block of consecutive data chosen thus arbitrarily from within a containing body of data, by simply evaluating it under the assumption of a background of an equal number of *independent* blocks, with a small corrective factor. Put more mathematically, if r consecutive runs are chosen from within data of S runs, $S - r + 1$ blocks of r consecutive runs can be formed. Then the observation is that possibly assuming the $S - r + 1$ blocks as uncorrelated will for practical purposes produce the same probability except for a known factor. There is some analogy here to the variance of the pure matching case. In that situation there is known correlation between the various positions of matching cards (2) and none in the binomial case. Yet to get the variance of the matching case from that of the corresponding binomial, one need only apply the simple corrective factor $n/(n-1)$. Since there is a proper accepted method furnished by Tippett of evaluating in the case of a background of *independent* blocks of data (4, 7) this sort of evaluation would be desirable.

DISCUSSION

Until the theoretical background of these empirically suggested problems is solved exactly, or sufficiently approximated, no unchangeable set of rules of proper procedure can be laid down for the *internal* interpretation of ESP data.

In the evaluation of a given subject's data there is a correction which in some instances should be applied to the probability obtained by direct recourse to the tables. This correction, as has been shown in this paper, is required by the fact that the stopping point may be in some instances and in some measure a function of the results obtained up to that point—i.e., optional stopping.

[2] Estimated from Table IV, reference 6.

There is further revision required for the independent evaluation of any special block of data (e.g. those of a single subject). This is the revision mentioned earlier as suggested by Tippett, correcting for the sampling from a larger background. Concerning Tippett's method and its extension to several blocks of data of differing probabilities, there will be further discussion in a later publication.

The results call quite clearly for some revision of the level of significance of critical ratio under the optional stopping procedures assumed for the basis of this paper. A probability criterion for significance of 1 in 150 may be met by $X = 3.1$ when the number of runs is less than 100, and $X = 3.3$ would probably be approximately equivalent when the number of runs involves several hundred. The value $X = 3.1$ is obtained by Table IV of reference 6, extended, taking into account the factor by which the empirical probability is larger than the desired one. At $X = 3.3$ the probability is halved again which is interpreted in the empirical series results as an expected doubling of the number of successful blocks from 100 runs up to several hundred.

These estimates do not, of course, apply to a series at the outset of which the exact number of runs to be undertaken was specified, or in which the stopping point was determined *clearly independently of the results.* In such cases the usual procedure of finding a probability would apply, subject, of course to its final interpretation against the legitimate background of, say, all investigations made under effectively similar conditions which are available for consideration.

It is not, however, a *requirement* that the data from all experimental conditions designed to bear upon a hypothesis in question be pooled and evaluated together—only those of essentially similar conditions need be pooled. On the other hand, results may be pooled (in their entirety) when the experimental conditions are sufficiently similar to warrant generalization. Differences in conditions which conceivably may affect the hypothetical principle under investigation warrant independent grouping and evaluation. Only thus can statistical method be an effective instrument of inquiry.

This paper has been concerned mainly with problems of the ESP research, but its bearing is by no means thus confined and should be applicable to any research in which the point of discontinuance is optional and may be affected by knowledge of the results. On the other hand, its bearing upon the ESP research itself should be clearly limited in application. It is evident, first, from the reports that certain work is exempt from the bearing of the corrections for optional stopping suggested by this paper; and it is recognized that these refinements involve

only marginal probabilities, not in any instance crucial to the general research. The large critical ratios obtained from the ESP investigation can well undergo such modifications as would at most be required, without serious consequences to their significance.

REFERENCES

1. GREENWOOD, J. A. Analysis of a large chance control series of ESP data. *J. Parapsychol.*, 1938, **2**, 138-146.
2. GREENWOOD, J. A. Variance of a general matching problem. *Annals of Mathematical Statistics*, 1938, **9**, No. 1, 56-59.
3. GREENWOOD, J. A. AND STUART, C. E. Mathematical techniques used in ESP research. *J. Parapsychol.*, 1937, **1**, 206-225.
4. HOTELLING, H. AND FRANKEL, F. R. The transformation of statistics to simplify their distribution. *Annals of Mathematical Statistics*, 1938, **9**, No. 2, see p. 88.
5. HUNTINGTON, E. V. A rating table for card-matching experiments. *J. Parapsychol.*, 1937, **1**, 292-294.
6. STUART, C. E. AND GREENWOOD, J. A. A review of criticisms of the mathematical evaluation of ESP data. *J. Parapsychol.*, 1937, **1**, see p. 303.
7. TIPPETT, L. N. C. The methods of statistics. London: Williams & Northgate, Ltd., 1931. Page 53.

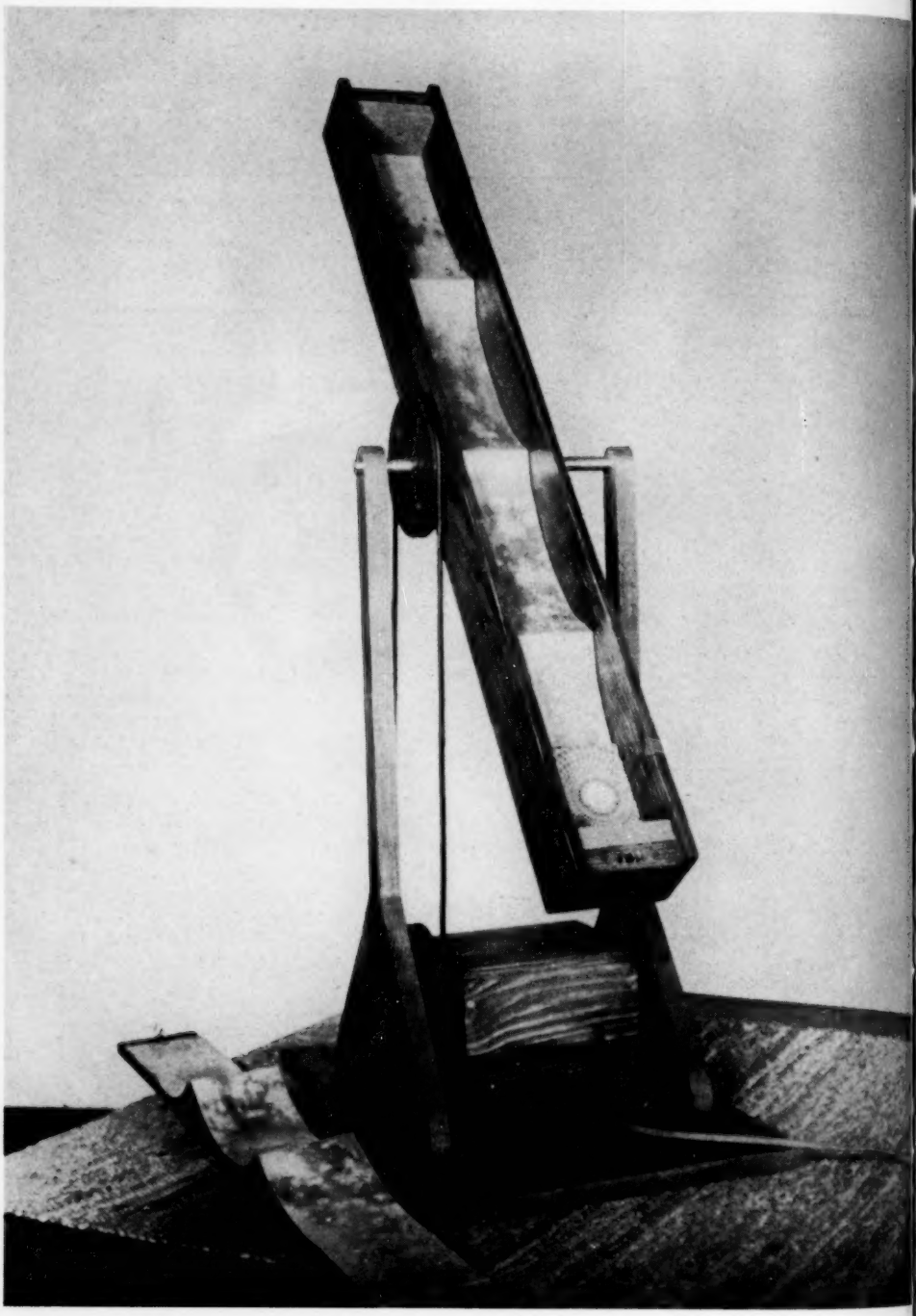


FIG. 1. SHUFFLING MACHINE

RESEARCH NOTES

NOTE ON A SHUFFLING MACHINE

It is occasionally feasible to use mechanical shuffling of the cards used in ESP tests. Since random re-ordering of the cards is required, such machines as are on the market for Bridge dealing are not wholly satisfactory, because the order after dealing is a strict and computable, though complex, function of the order of the cards when placed in the machine. A device designed by Mr. O. K. Merritt and constructed in the Duke University Parapsychology Laboratory is intended to approximate mechanically a sliding hand shuffle.

The shuffling machine (see illustration opposite) consists of three essential parts: the elongated box in which the cards are shuffled; the motor mechanism; and the supporting frame. The box, whose outside and inside dimensions respectively are 2.75 x 3.75 x 31.75 inches and 2 x 3 x 29 inches, is made of ⅜ inch stock. The floor and roof of the box are lined with 3" sheet zinc bent in waves over which the cards slide as the box revolves. The speed of the falling cards is variable, depending upon the friction between cards and the zinc and the angle of the box at any given moment. At each end of the box there is a ⅝" sponge rubber cushion to prevent the ends of the cards from becoming battered. The box is so mounted that it makes complete revolutions, the cards sliding first along the bottom and returning via the top. The moving mechanism consists of an electrically driven motor and a rheostat by means of which the speed of the motor may be varied. The speed of the box may be adjusted between two and four revolutions per minute. The motor is connected to the box by means of gears, pulley, and belt. The base and supporting arms are made of ¾" stock, 11.5 x 18 x 21.75 inches.

A test of the adequacy of shuffling with this machine is herewith presented. A pack of new ESP cards was taken and the cards numbered consecutively from 1 to 25. The cards were placed in the machine and shuffled a predetermined number of revolutions of the shuffling box. The card order was recorded after each shuffle and compared with the preceding order and returned to the machine in that order for the next shuffle. The tests were made in groups of ten runs and the cards were always arranged in the sequence 1-25 for the first runs. This was done so that any natural tendency for like cards (cards of each symbol were numbered consecutively, e.g. circles were 1-5, crosses 6-10, etc.) to remain together would be given greater than chance opportunity to do so. One, three, and five revolutions of the shuffling box were chosen arbitrarily for the test, giving shuffling equivalent to two, four, and six sliding shuffles. Twenty runs were made with each degree of shuffling. The results are presented in the following tables:

TABLE 1

CORRESPONDENCE OF NUMBERED CARDS

Revolutions	Runs	Hits	av/25	Deviation
1.....	20	17	0.85	—3
3.....	20	22	1.10	2
5.....	20	21	1.05	1
Totals.....	60	60	1.00	00

TABLE 2

CORRESPONDENCE OF ESP SYMBOLS

Revolutions	Runs	Hits	av/25	Deviation
1	20	103	5.15	3
3	20	97	4.85	—3
5	20	104	5.20	4
Totals	60	304	5.06	4

It can easily be seen that these deviations are well within the range to be expected by chance. Even with a single revolution of the machine the cards are completely displaced from their positions relative to the whole pack so that it would be impossible to know the position of any card by a rational process. However, when only one revolution was given the cards, there was a greater tendency for groups of cards to remain together though they were shifted in relative position. The larger numbers of shuffles seem adequate to rearrange the cards with respect to each other and with respect to the pack as a whole.

The number of "doubles" (i.e. the number of times a card bearing one symbol was followed by a card with a like symbol) resulted in the following:

TABLE 3

Revolutions	Runs	Doubles	Average/25
1	20	142	7.1
3	20	80	4.0
5	20	101	5.0

In every group there was a strikingly large number of "doubles" in the first run and a very definite decline for the next few succeeding runs. This result is not unexpected, however, since the cards were always arranged with the maximum number of doubles (i.e., 20) at the beginning of each series of 10 runs. In ordinary use, the pack would be in a random order at the beginning and no great excess of doubles would be likely to occur.

BURKE M. SMITH.

Duke University.

A SUMMARY OF SOME NEGATIVE EXPERIMENTS

This report is based on forty thousand card guessing experiments, involving seventy different subjects, designed to test clairvoyance.* The cards used were the familiar ESP decks, and both the DT and the BT methods were employed. The results were negative. No individual scored beyond chance expectations nor did the average of the combined scores exceed normal expectancy. General procedures were as follows:

SUBJECTS. Sixty-five of the seventy subjects used in controlled tests were male college students ranging in age from eighteen to twenty-eight years. One, who did about a fourth of all the tests, was blind. The other five subjects were col-

* For help, both in carrying out these experiments and in summarizing these data, I am indebted to T. P. deGraffenried of Colgate University, who for two years has been a capable student assistant.

lege faculty members of whom the experimenter was one. The students were selected from a group of several hundred who had been given preliminary tests. Those chosen for the controlled work were the ones whose scores deviated most from an average of five hits per twenty-five trials in the preliminary work. All subjects had expressed an interest in the work and none was paid for his time. In about half of the cases a slight reward was offered for high scores—a steak dinner to anyone averaging higher than 6.5 for two hundred and fifty guesses.

The amount of work done by the various subjects is summarized in the following table:

Number of Subjects	Tests for each Subject	Total
51	250	12,750
9	500	4,500
5	1,000	5,000
2	2,000	4,000
1	1,250	1,250
1	2,750	2,750
1	9,750	9,750
70	.....	40,000

For the purposes of this report the results will be summarized under three heads, as follows: (1) The Mike S. Series—results obtained from the blind subject who did 9,750 tests and was the best scorer of all the subjects; (2) the General Series—results obtained from the first sixty-nine subjects listed above; (3) the Combined Series—total results. In each case the results will be compared with a chance series constructed from the same record sheets as the attempted ESP series involved.

METHODS. As subjects reported for experimental work they were told something of the nature of the tests and were advised of the large amount of successful work already reported in the field. Everything possible was done to induce a "confident frame of mind." The work was about evenly divided between the DT and the BT methods. Where the latter was employed a different deck of cards was used for each run through the pack in order to avoid detection of individual differences. Detectable differences existed in all packs used. After twenty-five cards had been guessed they were turned up and scored *by the experimenter*—not called back by the subject as is too often done. (See Carpenter-Phalen tests, *Journal of Parapsychology*, Vol. 1, p. 39, for one type of difficulty involved in having subjects assist.)

THE NORMAL SERIES. For the purpose of comparing them with the called card groups, empirical chance curves were constructed for each group by comparing the call cards with actual cards for which they were not intended. As many of these chance combinations exceeded "chance expectancy" as did called combinations. When an individual scores, on at least two hundred or more guesses, two and a half times the standard deviation (which is approximately four times the probable error) some authors are inclined to call the score "significant"—although all it may signify is that enough calls have been made to warrant some excessive deviations. (A critical ratio, X (Dev./S.D.) of 2.5 for any individual is meaningless unless the individual score is plotted along with *all* other scores.)

RESULTS. (1) *The Mike S. Series:* Mike, who has been blind since infancy, is a brilliant student with a strong will to believe in ESP. His extreme motivation might be attributed to his blindness since ESP offered, perhaps, a new set of "eyes." He was elated by high scores, depressed by low ones. Most of his

work was done in blocks of two hundred and fifty guesses and at times when Mike felt most "psychic."

A. A comparison of "hits per 25" by attempted ESP and by chance combinations on 390 sets of 25; 9,750 trials.

Hits per 25	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
Mike's Scores.....	2	6	22	52	79	61	80	41	28	10	6	3	0	390
Chance Combinations.....	1	18	38	55	70	73	59	43	20	8	3	0	2	390

B. *Averages:* Mike's average number of hits per 25..... 5.13
Average by chance combinations..... 4.73

C. *"Significant variations":* (A summary of all instances of 100 trials or more where the critical ratio, X (Dev./S.D.), attained a value of 2.50 or more; considered by some writers to signify the operation of some principle other than chance.)

In the *ESP trials* of the Mike Series positive deviations (in which successive trials of 100 or more provide an X value of 2.50 or more) occurred four times. No "significant" negative deviations occurred.

Positive deviations: On one run of 250 trials Mike made 68 hits, giving X a value of 2.84; 66 of these hits were made in the first 225 trials of this series. Computing the critical ratio on the basis of 66 hits for 225 trials X has a value of 3.50.

On another sheet of Mike's scores we find an instance where 225 trials yielded 60 hits, giving X a value of 2.50.

On one run of 100 trials Mike scored 31 hits, giving X a value of 2.75; and on another he scored 30 hits for an X value of 2.50.

Negative deviations: None.

In the *chance trials* of the Mike Series positive deviations occurred three times; negative deviations occurred three times.

Positive deviations: One run of 250 "trials" yielded 67 hits for an X value of 2.68. One run of 100 "trials" yielded 31 hits for an X value of 2.75; another provided 30 hits for an X value of 2.50.

Negative deviations: One run of 250 "trials" provided only 32 hits, giving X a value of 2.84. One run of 200 "trials" provided only 22 hits, giving X a value of 3.18. One run of 100 "trials" provided only 9 hits, giving X a value of 2.75.

(2) *The General Series* (30,250 trials made by 69 subjects):

A. A comparison of "hits per 25" by attempted ESP and by chance combinations on 1,210 sets of 25, 30,250 trials.

Hits per 25	0	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
ESP Subjects.....	4	36	92	149	231	238	188	135	78	40	13	5	1	0	1,210
Chance Combinations..	4	29	85	167	252	255	163	125	78	32	14	4	1	1	1,210

B. *Averages:* ESP subjects average number of hits per 25..... 4.99
Average per 25 by chance combinations..... 4.92

C. *"Significant variations":* A summary of all instances (of 100 successive trials and more) where the critical ratio, X, attained a value of 2.50 or more.

In the *ESP trials* of the General Series positive deviations (in which successive trials of 100 or more provide an X value of 2.50 or more) occurred six times; negative deviations occurred three times.

Positive deviations (in ESP series): One run of 250 showed 66 hits—giving an X value of 2.53; 65 of these hits were made during the first 225 trials, and computing X on this basis we find it has a value of 3.33. Another run of 225 yielded 63 hits, for an X value of 3.00. Two runs of 100 showed 30 hits each, providing instances of X having a value of 2.50. One run of 100 showed 31 hits, giving X a value of 2.75. One run of 100 showed 33 hits, giving X a value of 3.25.

Negative deviations: One run of 250 trials showed only 28 hits—X equals 3.48. One run of 150 trials showed only 16 hits—X equals 2.88. One run of 100 trials showed only 10 hits—X equals 2.50.

In the *chance trials* of the General Series positive deviations occurred seven times; negative deviations occurred five times.

Positive deviations (in chance series): One run of 250 "trials" showed 67 "hits"—X equals 2.68; 46 of these hits were made during the first 150 "trials," and computing X on this basis we find it has a value of 3.27. Another run of 250 yielded 66 "hits" for an X value of 2.53. Three runs of 100 showed 30 "hits" each—X equals 2.50. Two runs of 100 showed 31 "hits" each—X equals 2.75.

Negative deviations: One run of 250 "trials" showed only 33 "hits"—X equals 2.68. The first 175 "trials" of this series showed only 20 "hits," and computing X on this basis we find that it equals 2.83. One run of 200 yielded only 23 "hits" —X equals 3.00. One run of 125 yielded only 13 "hits"—X equals 2.68. Two runs of 100 yielded only 10 "hits" each—X equals 2.50.

(3) *The Combined Series* (40,000 trials made by 70 subjects):

A. A comparison of "hits per 25" by attempted ESP and by chance combinations on 1,600 sets of 25, 40,000 trials.

ESP Subjects		Chance Combinations	
Hits Per 25	No. of Times Occurring	Hits Per 25	No. of Times Occurring
0	6	0	5
1	42	1	47
2	114	2	123
3	201	3	222
4	310	4	322
5	299	5	328
6	268	6	222
7	176	7	168
8	106	8	98
9	50	9	40
10	19	10	17
11	8	11	4
12	1	12	3
13	0	13	1
	1,600		1,600

B. *Averages:* ESP subjects average number of hits per 25..... 5.02
Per 25 by chance combinations..... 4.87

C. *"Significant variations":* Summarizing the instances where scores deviated from np to an extent where the critical ratio reached 2.50 or more, we find that the ESP subjects had 10 positive deviations of such an extent, and 3 negative ones, for a total of 13. The chance series showed 10 positive deviations where X equalled 2.50 or more and 8 negative ones, for a total of 18.

CONCLUSIONS. 1. *Statistical:* Empirical chance curves indicate that a critical ratio of 2.50 in isolated runs of even 250 or more trials is not great enough to warrant dismissing the "chance hypothesis." However the slight statistical error

which this might introduce in some of the published material on ESP would certainly *not* be great enough to vitiate the tremendously high scores produced in a good deal of the published experimental work.

2. *Experimental:* (a) The experiments produced some attempts at fraud on the part of the subjects. Hence the present investigator would be inclined to discount all unwitnessed and loosely controlled tests—simply to be on the safe side.

(b) Nearly all cards have perceptible individual differences. Consequently one ought to discount experiments where the subjects are allowed to handle individual cards or see the backs of them, unless a new pack is being employed. Mike, the blind boy, was able to identify ten cards by their individual "feel"—after being told their designs only once!

(c) Of some interest are 400 scattered guesses about which the subject involved felt "absolutely sure." They yielded 78 hits, an average of 4.87. Although these are relatively few scores to make generalizations about it would seem to indicate that there is no high correlation between subjective certainty and success in attempted ESP.

(d) Published estimates that one person in five has ESP ability seem far too high in the light of these experiments. Here seventy subjects were tested, none of whom was able to demonstrate the ability in question. Unpublished reports from a number of other investigators, both in the United States and England seem to support this criticism. In all, I know of some eight hundred subjects tried with universally negative results. The experimenters in question are established university scientists.

(e) Since I have been criticized for asserting that "the conditions for the appearance of ESP have not been objectively described sufficiently to warrant calling it an established scientific fact," perhaps a word on this score would be in point. If, as I suppose, an "established scientific fact" involves the necessity of competent investigators arriving at the same results—after following a prescribed technique, it seems premature to call ESP by such a name. To assume that all people who get negative results are incompetent is obviously to beg the question. To assume even some of them competent is to admit that "the conditions for the appearance ESP have not been objectively described sufficiently to warrant calling it an established scientific fact." Those who would investigate new fields can afford to be modest in their claims—in the present case they might reasonably claim to be examining a "lively" hypothesis![*]

E. T. ADAMS.

Colgate University.

[*Readers of the *Journal* need not, of course, be told that the experimental precautions Dr. Adams suggests are general routine today and are even considerably exceeded. But there are two important questions raised by him which are still very much alive and which merit comment.

First, the sampling problem (generalized): To what groups of data and under what circumstances can the critical ratio measure be applied? Dr. Adams rightly grants that the problem is a mere academic one so far as the evidence for ESP goes, because of the large critical ratios for a given entire research. But there is anticipated a real need for refined "legislation" on this problem for the many exploratory researches now under way. For the final code of propriety we must turn of course to mathematicians of probability who have furnished us with these instruments of measure in the past. In this issue Dr. Greenwood advances an empirical attack on some of these problems.

The second point raised concerns the failure of some ESP experimenters to get positive results. It does seem to be well demonstrated now that there are successful and unsuccessful ESP experimenters. Clark and Sharp began the distinction; Pratt and Price clinched it. Now in this issue MacFarland adds a further and a finer point. The will to succeed is not enough! Practically, the problem is to find those who can make good experimenters, as is done in any professional field. Theoretically, the problem is to find out what makes for failure and success as an experimenter.—EDITORS.]

GLOSSARY

In order to avoid constant redefining of commonly recurring terms in papers appearing in this journal, the following definitions are submitted for convenient reference. In case of any discrepancy between Glossary and usage in the text of an article the latter should be followed. Words defined elsewhere in the Glossary are italicized in the text of the definitions.

AGENT: In tests for *telepathy*, the person whose mental states are to be apprehended by the *percipient*.

AVERAGE SCORE = AVERAGE: Average number of *hits* per *run*.

BT (BEFORE TOUCHING): The technique in which the top card of the face-down *deck* is *called* and, after being *called*, is laid aside for *checking* at the end of the *run*. Each card in the *deck* is treated in the same way.

BM (BLIND MATCHING): The technique in which the *subject matches* a *deck* of *ESP cards* to five *key cards* which are laid out face down before him in an unknown order. Unless otherwise stated, the order is also unknown to the experimenter.

CALL v.: To designate a card by naming one suit either orally or by indicating it manually (as pointing or writing).

CALL n.: The response described above; also, the resulting selection.

CHANCE: The complex of undefined causal factors irrelevant to the purpose at hand.

CHANCE EXPECTATION = MEAN CHANCE EXPECTATION: The most likely *score* if only *chance* obtains.

CHANCE AVERAGE: *Mean chance expectation* in terms of *average* per *run*.

CHECK: To determine a *score* after the completion of a *run* by comparing the order of the *subject's calls* with the order of cards in the *deck*.

CLAIRVOYANCE: *Extra-sensory perception* of objective events as distinguished from *telepathic* perception (of the mental or subjective events of another person).

CRITICAL RATIO = X VALUE: The observed *deviation* divided by the *standard deviation*.

DECK: Twenty-five *ESP cards*, five of each suit.

DEVIATION: The amount an observed number of *hits* or an *average score* varies from the *mean chance expectation* or *chance average*. A *deviation* may be total (for a series of *runs*) or average (per *run*).

DT (DOWN THROUGH): The technique in which the cards are *called* down through the *deck* before any are removed or *checked*.

EMPIRICAL CONTROL: An experiment which wholly or partially follows the main experiment with the exception that the conditions are designed to exclude the possibility of *ESP*.

ESP (EXTRA-SENSORY PERCEPTION): Response to an external event (perception) not presented to any known sense.

ESP CARDS: *Cards*, each bearing one of the following five symbols: Star, circle, three parallel wavy lines (called "waves"), rectangle, plus.

ESP SYMBOLS: See plate opposite page 1, this journal Vol. I, No. 1, March, 1937.

ESP TESTS: A considerable number of techniques come under this heading, conveniently represented by initials, the principal ones being: *BT, DT, PT, GESP, BM, OM, STM.*

EXTRA-CHANCE: *Significantly* different from *chance expectation*.

GESP = GENERAL (OR UNDIFFERENTIATED) EXTRA-SENSORY PERCEPTION: A technique designed to test the occurrence of *extra-sensory perception*, permitting either *telepathy* or *clairvoyance* or both to operate.

KEY CARD: One of the five cards (where there are five suits) against which the cards of the test *deck* (i.e. *target cards*) in the *matching* tests (*OM, BM, TM*, etc.) are *matched.*

MATCHING: A form of *calling* in which a *target card* is placed opposite the *key card* which the *subject* selects to identify it.

OM (OPEN MATCHING): The technique in which the *subject matches* a *deck* of *ESP cards* to five *key cards* which are face up before him.

PERCIPIENT: The person who perceives. In this journal the common use will pertain to *ESP* rather than to sensory perception.

PT (PURE TELEPATHY) = TELEPATHY: The word "pure" emphasizes the exclusion of *clairvoyance.*

RELIABILITY: *Significance.*

RUN: The presentation and *calling* of 25 *ESP cards* or *symbols* in succession.

SCORE: The number of *hits* made in one *run.*

TOTAL SCORE: *Score* of any number of *runs.*

AVERAGE SCORE: *Total score* divided by number of *runs.*

SCREEN: An opaque barrier used between the subject and the card or agent. The main types of screens will be illustrated in this journal on their first introduction in print.

STM (SCREENED TOUCH MATCHING): The *touch matching* technique with an upright *screen* preventing vision by the *subject* of the handling of the cards by the experimenter.

SERIES: Several *runs* that are grouped consecutively or according to some other principle.

SIGNIFICANCE: The unlikelihood that a given event (or series) will not on the average occur by chance more often than once in 150 such events (or series). [Arbitrarily taken to mean a *deviation* in the expected direction such that the *critical ratio* is 2.5 times the *standard deviation* (or four times the probable error) or greater.]

STANDARD DEVIATION: A unit of measurement in statistical method; that *deviation* above and below *mean chance expectation* which is expected to include about ⅔ the *chance scores.* For *ESP cards,* S. D. $= 2.04 \sqrt{\text{no. of runs}}$.

SUBJECT: The person who is experimented upon. Most commonly the *percipient* in *ESP,* though also the *agent* in *telepathy.*

TARGET CARD: The card which the *percipient* is attempting to perceive (i.e. to *call,* or otherwise indicate a knowledge of).

TARGET DECK: The *deck* of cards the order of which the *subject* is *calling.*

TELEPATHY: *Extra-sensory perception* of the mental activities of another person. It does not include the *clairvoyant* perception of objective events.

TOUCH MATCHING: The technique in which the *subject* indicates his *call* by tapping or touching one of the five *key cards* while the *experimenter* places the *target card* thus called in front of the *key card* indicated.

TRIAL: A single *call* of a presented *symbol.*

X VALUE = CRITICAL RATIO: The observed *deviation* divided by the *standard deviation* or by the *probable error.*

WILLIAM McDOUGALL